university of groningen

faculty of mathematics
and natural sciences

Stein's Method Applied to Preferential Attachment Graphs

Bachelor Project Mathematics

July 2016

Student: S.A. Donderwinkel

First supervisor: Dr. D. Rodrigues Valesin

Second supervisor: Prof. dr. A.C.D. van Enter

Stein's Method Applied to Preferential Attachment Graphs

Serte Donderwinkel[1]

June 2016

Abstract

In probability and statistics, distributional limit theorems are used frequently. Stein's method is a method to quantify the error in distributional convergence and determine the distance between the distribution of different random variables. In this thesis, Stein's method is illustrated for different distributions. Up next, preferential attachment graphs are introduced. This is a type of random graphs in which vertices are added one by one and then connect to another vertex or itself with a probability proportional to the degree of the vertices. These graphs are also called 'rich-get-richer'-models. Stein's method for the negative binomial distribution is then applied to illustrate rates of convergence for the total variation distance between the distribution of the degree of a randomly chosen vertex and an appropriate distribution as the number of vertices tends to infinity. Furthermore, power-law behaviour in preferential attachment graphs is shown.

Keywords: Stein's method, negative binomial distribution, random graphs, dynamical networks, uniform attachment random graph model, preferential attachment random graph model

Contents

[1]University of Groningen. s.a.donderwinkel@student.rug.nl

1 Introduction

An important topic in statistics and probability is the distribution of a random variable approaching the distribution of another random variable, or in short, the study of convergence in distribution. Studying distributional limits is very valuable, since it identifies random variables with similar distributions and allows one to extend results for random variables with a specific distribution to 'similar' random variables. Thus, due to distributional limit theorems, we can use the results on extremely familiar distributions for an unlimited number of other distributions, simply because the random variables are close enough in an appropriate sense.

However, what is close enough? If we use distributional limit theorems to generalize results merely based on convergence, we should at least quantify how good this convergence is, right? That is where Stein's method comes in. Stein's method is a method to quantify the error in distributional convergence and determine for example the total variation distance between two distributions. This thesis gives an introduction to Stein's method, by applying it to a number of distributions. In Section 2 we demonstrate Stein's method for the standard normal distribution (Subsection 2.3), Poisson distribution (Subsection 2.4) and the Exponential distribution (Subsection 2.5). Up next, in Section 3 we apply Stein's method to the negative binomial distribution.

To illustrate how Stein's method works in practice, the rest of the thesis is devoted to obtaining results on preferential attachment graphs with use of Stein's method.
Preferential attachment graphs are a type of random graph in which new nodes are added one by one. A new node connects to one of the other nodes or to itself with probability proportional to the number of connections a certain node already has. The nodes with many connections are thus likely to get even more connections, and this is why these graphs are often called 'rich-get-richer'-graphs. This type of random graphs is a good approximation of some real-world dynamical networks, such as social networks, networks seen in biology and citation networks. A precise definition of preferential attachment graphs will be given in Subsection 5.1.
The random variable representing the number of connections, or the degree, of a randomly selected node from a preferential attachment graph can be approximated by a random variable having a mixture distribution of the negative binomial distribution. This convergence and the error in the approximation will be shown with Stein's method. This approximation also illustrates that power-law behaviour is seen in preferential attachment graphs. This means that when picking a random node from a preferential attachment graph, the probability that it has a certain degree decreases as a power law as the degree increases.

To get familiar with random graphs and applying Stein's method, we will start with Stein's method for the geometric distribution applied to the uniform attachment random graph model

in Section 4. In Subsection 4.1 we use our results for the negative binomial distribution to obtain results for the geometric distribution, which is a special case of the negative binomial distribution. Then, in Subsection 4.2, we apply these results to the uniform attachment random graph model. We will obtain an error bound on the distribution of the random variable representing the degree of a randomly selected node.
Now that we have applied Stein's method to a type of random graphs that is among the simplest of random graphs, we will apply it to preferential attachment graphs. We will start with a less general type of preferential attachment graphs in Section 5 and apply Stein's method for the geometric distribution to it.

Thereafter we will continue to the core of the thesis, which is applying Stein's method for the negative binomial distribution to preferential attachment graphs. This will be done in Section 6. Also the result on power-law behaviour in preferential attachment graphs is obtained in this section.

To summarize, the thesis will first cover Stein's method and then apply Stein's method to different types of random graphs. Chapter 6, which is the main result described in this thesis, is based on [Ross, 2013]. The other chapters will serve as support and provide the reader with context and pre-knowledge to Chapter 6.

2 Introduction to Stein's Method

An important topic in statistics and probability is proving distributional limit theorems. This section is based on [Reinert, 2011], [Pitman and Ross, 2012] and [Ross, 2011].

2.1 Some important basics

To be able to discuss distributional limit theorems properly, we will need to define *convergence in distribution*, the concept of events happening *almost surely* (or abbreviated, *a.s.*) and *probability metrics.*

Definition 2.1 *Suppose that $X_1, X_2, \dots$ and X are real-valued random variables with cumulative distribution functions $F_1, F_2, \dots$ and F respectively. Then, the distribution of the sequence X_n converges to the distribution of X as $n \to \infty$, or, equivalently, X_n converges to X in distribution as $n \to \infty$ if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all $x \in \mathbb{R}$ where F is continuous.

Note that for convergence in distribution to make sense, the random variables do not even need to be defined on the same probability space. In fact, the definition of the convergence only includes the distribution functions, and not the random variables.

Definition 2.2 *An event happens almost surely if it happens with probability 1. In other words, the set of possible exceptions may be non-empty, but it has probability 0.*

Note that this concept is identical to stating a statement is true almost everywhere in measure theory. Now, we will define probability metrics in order to have a notion of distance between two probability measures.

Definition 2.3 *Let μ and ν be two probability measures on a σ-algebra $\mathcal{F}$ in a sample space Ω. Let $\mathcal{H}$ be a family of test functions mapping Ω to $\mathbb{R}$. Then, we define the probability metric*

$$\mathrm{d}_{\mathcal{H}}(\mu,\nu) = \sup_{h\in\mathcal{H}} \left| \int_{\Omega} h(x)\mathrm{d}\mu(x) - \int_{\Omega} h(x)\mathrm{d}\nu(x) \right|.$$

For random variables X and Y with laws μ and ν, we will often write $\mathrm{d}_{\mathcal{H}}(X,Y)$ instead of $\mathrm{d}_{\mathcal{H}}(\mu,\nu)$ by abuse of notation. Then, we can write

$$d_{\mathcal{H}}(X,Y) = \sup_{h\in\mathcal{H}} |\mathrm{E}h(X) - \mathrm{E}h(Y)|.$$

By picking different sets of test functions $\mathcal{H}$, different metrics can be defined. A very important metric to be considered in this paper, is the total variation distance between distributions. It is defined as follows.

Definition 2.4 *Take $\mathcal{H} = \{\mathbb{I}[A \in \mathbb{R}] : A \in \mathrm{Borel}(\mathbb{R})\}$ in Definition 2.3. Then, we get a metric which we define to be the* total variation metric, *denoted by d_{TV}.*

The total variation metric is often used for discrete distributions. An important result is that for two integer-valued, non-negative random variables X, Y,

$$d_{TV}(X,Y) = \frac{1}{2} \sum_{k\in\mathbb{N}\cup\{0\}} |\mathrm{P}(X=k) - \mathrm{P}(Y=k)|. \tag{2.1}$$

2.2 Distributional limit theorems

Probably the best-known example of a distributional limit theorem is the central limit theorem, which is stated as follows.

Theorem 2.5 *Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with a distribution with expected value μ and finite variance σ^2. Then, considering the random variable*

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

and the random variable

$$Z_n = \frac{M_n - \mu}{\sigma/\sqrt{n}},$$

the distribution of Z_n converges to the standard normal distribution as $n \to \infty$.

Remember that the standard normal distribution is a distribution with mean 0 and variance 1, which has the probability density function

$$\phi(x) = \frac{1}{2\pi} e^{-x^2/2}$$

with $x \in \mathbb{R}$. Note that this theorem is very powerful, since the convergence of the distribution of Z_n to the standard normal distribution is independent of the distribution of X_i. However, as powerful as this statement is, it does not give any quantification of the convergence. Therefore, when taking a random sample, the error bound on the distance between Z_n and the normal distribution is not known. This is where Stein's method comes in.

2.3 Stein's method for the standard normal distribution

In Theorem 2.1, convergence in distribution is dependent on the cumulative distribution function, which characterizes the distribution. Stein's method however, does not make use of a characterizing function as a basis for distributional convergence, but makes use of *characterizing operators*. An example of such a characterizing operator for the standard normal distribution is given by Stein's Lemma.

Lemma 2.6 *Define the operator $\mathcal{A}$ by*

$$\mathcal{A}f(x) = f'(x) - xf(x).$$

Then, the following statement holds.

1. *If Z has the standard normal distribution,*

   $$\mathrm{E}\mathcal{A}f(Z) = 0$$

   for all absolutely continuous f with $\mathrm{E}|f'(Z)| < \infty$.

2. *If for some random variable W,*

   $$\mathrm{E}\mathcal{A}f(W) = 0$$

   for all absolutely continuous f with $\mathrm{E}|f'(Z)| < \infty$, then W has the standard normal distribution.

The operator $\mathcal{A}$ is thus referred to as a characterizing operator of the standard normal distribution. Note that the two statements are not symmetrical, since in the second statement the condition on f depends on the distribution of Z and not on the distribution of W.

In the proof of Stein's Lemma and to achieve later results, we will need the following lemma.

Lemma 2.7 *Let $\Phi(x)$ denote the cumulative distribution function of the standard normal distribution. Then, the differential equation*

$$\mathcal{A}f_x(w) = \mathbb{I}[w \leq x] - \Phi(x) \tag{2.2}$$

has a unique bounded solution f_x, which is given by

$$f_x(w) = e^{w^2/2} \int_w^\infty e^{-t^2/2}(\Phi(x) - \mathbb{I}[t \leq x])dt.$$

Proof. We will show this statement by the method of integrating factors. Use the integrating factor $e^{-w^2/2}$. Then we get that,

$$e^{-w^2/2}\left(f_x'(w) - wf_x(w)\right) = \frac{d}{dw}\left(e^{-w^2/2}f_x(w)\right) = e^{-w^2/2}\left(\mathbb{I}[w \leq x] - \Phi(x)\right).$$

Integrating both sides of the second equality sign and multiplying by $e^{w^2/2}$ yields

$$f_x(w) = e^{w^2/2} \int_w^\infty e^{-t^2/2}(\Phi(x) - \mathbb{I}[t \leq x])dt + Ce^{w^2/2},$$

as the general solution of the differential equation with arbitrary constant C. Now, what is left to show is that only one value of C yields a bounded solution.
To this end, we will show that the solution is bounded for $C = 0$. Then, obviously, it is not bounded for other values of C, since $e^{w^2/2}$ is not bounded.
We will first show that

$$\int_{-\infty}^{\infty} e^{-t^2/2}(\Phi(x) - \mathbb{I}[t \leq x])dt = 0,$$

from which follows that

$$f_x = e^{w^2/2} \int_{w}^{\infty} e^{-t^2/2}(\Phi(x) - \mathbb{I}[t \leq x])dt = -e^{w^2/2} \int_{-\infty}^{w} e^{-t^2/2}(\Phi(x) - \mathbb{I}[t \leq x])dt.$$

Indeed,

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-t^2/2}\Phi(x)dt &= \Phi(x) \int_{-\infty}^{\infty} e^{-t^2/2}dt \\ &= \sqrt{2}\Phi(x) \\ &= \int_{-\infty}^{x} e^{-t^2/2}dt \\ &= \int_{-\infty}^{\infty} I[t \leq x]e^{-t^2/2}dt,\end{aligned}$$

which proves the statement.

Now, consider f_x for $w \leq x$. Then,

$$\begin{aligned}f_x(w) &= -e^{w^2/2} \int_{-\infty}^{w} e^{-t^2/2}(\Phi(x) - \mathbb{I}[t \leq x])dt \\ &= -e^{w^2/2} \int_{-\infty}^{w} e^{-t^2/2}(\Phi(x) - 1)dt \\ &= e^{w^2/2}\sqrt{2\pi}\Phi(w)(1 - \Phi(x)).\end{aligned}$$

The last equality follows from

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}dt.$$

Similarly, consider f_x for $w > x$. Then,

$$\begin{aligned}f_x(w) &= e^{w^2/2} \int_{w}^{\infty} e^{-t^2/2}(\Phi(x) - \mathbb{I}[t \leq x])dt \\ &= e^{w^2/2} \int_{w}^{\infty} e^{-t^2/2}\Phi(x)dt \\ &= e^{w^2/2}\sqrt{2\pi}\Phi(x)(1 - \Phi(w)).\end{aligned}$$

Again, the last equality follows from the definition of $\Phi(x)$ as the c.d.f. of the standard normal distribution. Thus,

$$f_x(w) = \begin{cases} e^{w^2/2}\sqrt{2\pi}\Phi(w)(1-\Phi(x)), w \leq x \\ e^{w^2/2}\sqrt{2\pi}\Phi(x)(1-\Phi(w)), w > x \end{cases}$$

Now, we will start determining a bound on f_x. We will first show

$$1 - \Phi(w) \leq \frac{1}{2}e^{-w^2/2}, w > 0.$$

Thus, we will show that $g \geq 0$ for all $w > 0$, with

$$g(w) = \frac{1}{2}e^{-w^2/2} + \Phi(w) - 1$$

The statement will be shown using derivatives.
Note that

$$\lim_{w\to\infty} g(w) = \lim_{w\to\infty}\left(\frac{1}{2}e^{-w^2/2} + \Phi(w) - 1\right) = 0 + 1 - 1 = 0, \tag{2.3}$$

and

$$\lim_{w\to 0} g(w) = \lim_{w\to 0}\left(\frac{1}{2}e^{-w^2/2} + \Phi(w) - 1\right) = \frac{1}{2} + \frac{1}{2} - 1 = 0. \tag{2.4}$$

Furthermore, since the equality

$$0 = \frac{dg}{dw} = \frac{-w}{2}e^{-w^2/2} + \frac{1}{\sqrt{2\pi}}e^{-w^2/2} = e^{-w^2/2}\left(\frac{-w}{2} + \frac{1}{\sqrt{2\pi}}\right)$$

has merely one solution, g has merely one extremum, namely at $w = \sqrt{\frac{2}{\pi}}$. From this observation and equations (2.3) and (2.4), we see that all that is left to prove is that g has a maximum at $w = \sqrt{\frac{2}{\pi}}$. And indeed,

$$\frac{d^2g}{dw^2}\Big|_{w=\sqrt{\frac{2}{\pi}}} < 0,$$

and we can conclude that g is non-negative for $w > 0$. Thus, we can conclude that

$$1 - \Phi(w) \leq \frac{1}{2}e^{-w^2/2}, w > 0.$$

Thus, for $w \leq x$,

$$|f_x(w)| = |e^{w^2/2}\sqrt{2\pi}\Phi(w)(1-\Phi(x))| \leq |e^{w^2/2}\sqrt{2\pi}\Phi(w)\frac{1}{2}e^{-x^2/2}| \leq \sqrt{\frac{\pi}{2}}.$$

The last inequality follows from the fact that $0 \leq \Phi(w) \leq 1$ and $0 < e^{w^2/2 - x^2/2} \leq 1$, since $0 < w \leq x$.
Furthermore, for $w > x$,

$$|f_x(w)| = |e^{w^2/2}\sqrt{2\pi}\Phi(x)(1-\Phi(w))| \leq |e^{w^2/2}\sqrt{2\pi}\Phi(x)\frac{1}{2}e^{-w^2/2}| \leq \sqrt{\frac{\pi}{2}}.$$

Thus, $|f_x|$ is bounded for $C = 0$, and thus, this yields the unique solution f_x of the differential equation ∎

We will now prove Stein's Lemma, i.e. Lemma 2.6.

Proof. Let us first prove the first part of the statement. Assume the random variable Z has the standard normal distribution. Then,

$$\begin{aligned}
\mathrm{E}f'(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(t)e^{-t^2/2}dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} f'(t) \left[\int_{t}^{\infty} we^{-w^2/2}dw \right] dt + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} f'(t) \left[\int_{-\infty}^{t} we^{-w^2/2}dw \right] dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} we^{-w^2/2} \left[\int_{0}^{w} f'(t)dt \right] dw + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} we^{-w^2/2} \left[\int_{w}^{0} f'(t)dt \right] dw \\
&= \mathrm{E}Zf(Z)
\end{aligned}$$

Here we use that the probability density of the standard normal distribution is given by $\phi(w) = \frac{1}{\sqrt{2\pi}}e^{-w^2/2}$. Furthermore, for the second equality, we split up the interval of integration and we use that

$$\frac{d}{dt}\left(-e^{-t^2/2}\right) = te^{-t^2/2},$$

and thus

$$\int_{t}^{\infty} we^{-w^2/2}dw = \lim_{s\to\infty} -e^{-w^2/2}|_{w=t}^{w=s} = e^{-t^2/2}.$$

Similarly,

$$\int_{-\infty}^{t} we^{-w^2/2}dw = -e^{-t^2/2}.$$

Finally, we switch the order of integration in the third equality. Thus,

$$\mathrm{E}\mathcal{A}f(Z) = 0.$$

To show the second part of the statement, assume that W is a random variable such that $\mathrm{E}\mathcal{A}f(W) = 0$ for all bounded, continuous and piecewise continuously differentiable functions f with $\mathrm{E}|f'(Z)| < \infty$. The solution f_x to the differential equation (2.2) is such a function. Thus, for all x,

$$0 = \mathrm{E}\mathcal{A}f_x(W) = \mathrm{E}[\mathbb{I}[W \leq x] - \Phi(x)] = P[W \leq x] - \Phi(x)$$

Thus, $P[W \leq x] = \Phi(x)$, and since the c.d.f. uniquely characterizes a distribution, W has the standard normal distribution. ∎

Now we will show how this characterizing operator for the standard normal distribution can be used to quantify the distance between two distributions.

Corollary 2.8 *Let $\Phi(x)$ denote the distribution function of the standard normal distribution. Let $f_x(w)$ denote the unique bounded solution of the differential equation (2.2). Then, for any random variable W,*

$$|\mathrm{P}(W \leq x) - \Phi(x)| = |\mathrm{E}\mathcal{A}f_x(W)|. \tag{2.5}$$

Equation (2.5) is obviously obtained by inserting the random variable W in Equation (2.2) and then successively taking the expected value and the absolute value on both sides of the

equality sign.
Corollary 2.8 suggests a method for bounding the difference between a random variable with the normal distribution and another random variable. If we find a bound on

$$|\mathrm{E}[f_x'(W) - Wf_x(W)]|,$$

we also bound the difference between the c.d.f. of W and the c.d.f. of the standard normal distribution.
Though a bound on the difference between the c.d.f.'s might give us useful information on the distance between two distributions, it would be better if we found a bound on one of the earlier defined metrics between two distributions. Recall that for two random variables X and Y and some set of test functions $\mathcal{H}$, the metric was given by

$$d_{\mathcal{H}}(X,Y) = \sup_{h\in\mathcal{H}}[\mathrm{E}h(X) - \mathrm{E}h(Y)]$$

Note that if we denote

$$g(w) = \mathbb{I}[w \leq x]$$

and consider this function as a test function, we have shown in Corollary 2.8 that

$$|\mathrm{E}g(X) - \mathrm{E}g(Z)| \leq |\mathrm{E}\mathcal{A}f_x(X)| \tag{2.6}$$

with X any random variable and Z being a random variable with the standard normal distribution. Here, f_x is the solution of the differential equation

$$\mathcal{A}f(w) = g(w) - \mathrm{E}g(Z),$$

where Z is again a random variable with the standard normal distribution.
Thus, we have already obtained a bound on

$$[\mathrm{E}h(X) - \mathrm{E}h(Z)]$$

for a particular test function. This bound depends on a differential equation depending on the test function. In [Chen et al., 2010] it is shown that analogous results to (2.6) hold for different test functions. Let h denote a test function. If f_h denotes a solution of the differential equation (2.2).

$$\mathcal{A}f(w) = h(w) - \mathrm{E}h(Z),$$

then

$$|\mathrm{E}h(X) - \mathrm{E}h(Z)| = |\mathrm{E}\mathcal{A}f_h(X)|. \tag{2.7}$$

Equation (2.7) is used in proving the following lemma.

Lemma 2.9 *If W is a random variable and Z has the standard normal distribution, then*

$$d_{\mathcal{H}}(W,Z) = \sup_{h\in\mathcal{H}} \mathrm{E}\mathcal{A}f_h(W).$$

This result is an example of the general setup of Stein's method.

2.4 Stein's method for the Poisson distribution

In the last section we have shown how Stein's method can quantify the difference between the distribution of a random variable and the standard normal distribution. We first found a characterizing operator. Then we needed to solve a differential equation. This yielded a bound on the error in the c.d.f.s of the two random variables. By generalization, a bound on the metric of the two distributions could be found.
In this section, we will follow the same outline to apply Stein's method to the Poisson distribution.
Recall that the Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed time if these events occur with a known average rate λ and independently of the time since the last event. Its probability mass function is given by

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

for $k = 0, 1, ...$ Firstly, we will define the operator $\mathcal{B}$, which is a characterizing operator for the Poisson distribution.

Lemma 2.10 *For $\lambda > 0$, define the functional operator $\mathcal{B}$ by*

$$\mathcal{B}f(k) = \lambda f(k+1) - kf(k).$$

Then, the following statements hold.

1. *If the random variable Z has the Poisson distribution with mean λ, then $\mathrm{E}\mathcal{B}f(Z) = 0$ for all bounded functions f.*
2. *If for some non-negative integer-valued random variable W, $\mathrm{E}\mathcal{B}f(W) = 0$ for all bounded functions f, W has the Poisson distribution with mean λ.*

The operator $\mathcal{B}$ is referred to as the characterizing operator of the Poisson distribution.

When we proved Stein's lemma, our next step was to solve a differential equation. Now, since we have to do with a discrete random variable, our next step is to solve a difference equation.

Lemma 2.11 *Let P_λ denote the probability with respect to a Poisson distribution with mean λ and let $A \subseteq \mathbb{N} \cup \{0\}$. Then, there exists a unique solution f_A of*

$$\mathcal{B}f(k) = \mathbb{I}[k \in A] - P_\lambda(A) \tag{2.8}$$

with $f_A(0) = 0$. It is given by

$$f_A(k) = \lambda^{-k} e^{\lambda} (k-1)! \left[P_\lambda(A \cap U_k) - P_\lambda(A) P_\lambda(U_k) \right],$$

where $U_k = \{0, 1, ..., k-1\}$.

Proof. Since the solution is defined recursively, the solution is obviously unique under the initial condition $f_A(0) = 0$.
We insert the conjectured solution in the definition of $\mathcal{B}f(k)$ to show that Equation (2.8) holds for this function f_A.

$$\begin{aligned}\mathcal{B}f_A(k) =& \lambda f_A(k+1) - kf_A(k) \\ =& \lambda\lambda^{-k-1}e^{\lambda}k!\left[P_\lambda(A\cap U_{k+1}) - P_\lambda(A)P_\lambda(U_{k+1})\right] \\ & - k\lambda^{-k}e^{\lambda}(k-1)!\left[P_\lambda(A\cap U_k) - P_\lambda(A)P_\lambda(U_k)\right] \\ =& \lambda^{-k}e^{\lambda}k!\left[(P_\lambda(A\cap U_{k+1}) - P_\lambda(A\cap U_k)) - (P_\lambda(A)P_\lambda(U_{k+1}) - P_\lambda(A)P_\lambda(U_k))\right] \\ =& \lambda^{-k}e^{\lambda}k!\left[P_\lambda(\{k\})\mathbb{I}[k\in A] - P_\lambda(A)P_\lambda(\{k\})\right]\end{aligned}$$

$= \mathbb{I}[k \in A] - P_\lambda(A)$

The last equality holds, because

$$P_\lambda(\{k\}) = \frac{\lambda^k e^{-\lambda}}{k!} = \left(\lambda^{-k}e^{\lambda}k!\right)^{-1}.$$

This proves that f_A solves the difference equation. ∎

We will now prove Lemma 2.10.

Proof. We will first prove the first statement. Assume the random variable Z has the Poisson distribution with mean λ. Let f be a bounded function. We will show that

$$\lambda\mathrm{E}[f(Z+1)] = \mathrm{E}[Zf(Z)].$$

Indeed,

$$\begin{aligned}\lambda\mathrm{E}[f(Z+1)] &= \lambda\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}f(k+1) \\ &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{k+1}}{(k+1)!}f(k+1)(k+1) \\ &= \mathrm{E}[Zf(Z)]\end{aligned}$$

Now, for the second statement, assume W is a random variable such that $\mathrm{E}\mathcal{B}f(W) = 0$ for all bounded functions f. Take $j \in \mathbb{N}\cup\{0\}$ and take $f_j(k) = \mathbb{I}[k=j]$. Then, $\mathrm{E}\mathcal{B}f_j(W) = 0$ implies that

$$0 = \mathrm{E}\left[\lambda\mathbb{I}[j = W+1] - W\mathbb{I}[j=W]\right] = \lambda\mathrm{P}(W = j-1) - j\mathrm{P}(W=j).$$

The relation

$$P(W=j) = \frac{\lambda}{j}P(W=j-1)$$

identifies W as a random variable with the Poisson distribution. ∎

These results allow us to prove the main result of this subsection.

Corollary 2.12 *If W is an integer-valued, non-negative random variable with mean λ, then*

$$|\mathrm{P}(W\in A) - P_\lambda(A)| = |\mathrm{E}[\lambda f_A(W+1) - Wf_A(W)]|. \tag{2.9}$$

Equation (2.9) is obviously obtained by inserting the random variable W in Equation (2.8) and then sequentially taking the expected value and the absolute value on both sides of the equality sign.

We also have an analogue of Lemma 2.9 for the Poisson distribution, which is as follows.

Lemma 2.13 *Let W be an integer-valued, non-negative random variable with mean λ and let $\mathcal{F}$ be the set of functions f such that*

$$|f_A(k)| \leq \min\{1, \lambda^{-1/2}\} \text{ and } |f(k+1) - f(k)| \leq \min\{1, \lambda^{-1}\}$$

for all k. Then, if Z is a random variable with the Poisson distribution with mean λ,

$$d_{TV}(W, Z) \leq \sup_{f \in \mathcal{F}} |\mathrm{E}\mathcal{B}f(W)|.$$

This result is proven in [Barbour et al., 1992].

We have now obtained similar results for the Poisson distribution as we have previously obtained for the normal distribution.

2.5 Stein's method for the exponential distribution

Our next example is Stein's method for the exponential distribution. As is to be expected, the outline of the method is similar to that for the normal distribution and Poisson distribution. Recall that the exponential distribution is the distribution that describes the time between events in a Poisson process. A random variable has the exponential distribution with rate λ if its density is given by

$$f(z) = \lambda e^{-\lambda z}$$

for $z > 0$. It has mean λ^{-1} and variance λ^{-2}. We will now define a characterizing operator for the exponential distribution.

Lemma 2.14 *Define the operator $\mathcal{C}$ by*

$$\mathcal{C}f(x) = f'(x) - f(x) + f(0)$$

Then, the following statements hold.

1. *If the random variable Z has the exponential distribution with rate 1, then $\mathrm{E}\mathcal{C}f(Z) = 0$ for all absolutely continuous functions f with $\mathrm{E}|f'(Z)| < \infty$.*
2. *Let Z be a random variable with the normal distribution. If for some non-negative random variable W, $\mathrm{E}\mathcal{C}f(W) = 0$ for all absolutely continuous functions f with $\mathrm{E}|f'(Z)| < \infty$, then W has the exponential distribution with rate 1.*

The operator $\mathcal{C}$ is referred to as the characterizing operator of the exponential distribution.

Again, our next step is to solve a differential equation, which is given in the following lemma.

Lemma 2.15 *Let $E(x)$ denote the cumulative distribution function of the exponential distribution with rate 1. Then, the differential equation*

$$\mathcal{C}f_x(w) = \mathbb{I}[0 < w \leq x] - E(x) \tag{2.10}$$

with $x > 0$ has a unique solution f_x such that $f_x(0) = 0$. It is given by

$$f_x(w) = -e^w \int_w^\infty e^{-t}(\mathbb{I}[t \leq x] - E(x))dt.$$

Proof. Equation (2.10) can be written as

$$\mathbb{I}[w \leq x] - E(x) = f_x'(w) - f_x(w) + f_x(0) = f_x'(w) - f_x(w)$$

Using the integrating factor e^{-w}, we obtain

$$e^{-w}(f_x'(w) - f_x(w)) = \frac{d}{dw}\left(e^{-w} f_x(w)\right) = e^{-w}(\mathbb{I}[w \leq x] - E(x))$$

Integrating w.r.t. w and multiplying by e^{w} yields

$$f_x(w) = e^{w} \int_0^w e^{-t}(\mathbb{I}[0 < t \leq x] - E(x))dt + C$$

for arbitrary constant C. Obviously, for $f_x(0) = 0$ to hold, C has to equal 0.
Now, all that's left to show is that

$$e^{w} \int_0^w e^{-t}(\mathbb{I}[0 < t \leq x] - E(x))dt = -e^{w} \int_w^\infty e^{-t}(\mathbb{I}[0 < t \leq x] - E(x))dt,$$

or, equivalently, that

$$\int_0^\infty e^{-t}(\mathbb{I}[0 < t \leq x] - E(x))dt = 0.$$

Indeed,

$$\begin{aligned}
\int_0^\infty e^{-t} E(x) dt &= E(x) \int_0^\infty e^{-t} dt \\
&= E(x) \\
&= \int_0^x e^{-t} dt \\
&= \int_0^\infty e^{-t} \mathbb{I}[0 < t \leq x] dt.
\end{aligned}$$

The second equality holds, since e^{-t} is the density function of the exponential distribution and it is integrated over its domain. This proves the statement and shows that the conjectured solution indeed solves the differential equation and is unique. ∎

Now, we are ready to prove Lemma 2.14.
Proof. To show statement 1, let Z be a random variable with the exponential distribution

with rate 1. Let f be an absolutely continuous function with $\mathrm{E}|f'(Z)| < \infty$.

$$\begin{aligned}
\mathrm{E}[f'(Z)] &= \int_0^\infty f'(t)e^{-t}dt \\
&= \int_0^\infty f'(t) \int_t^\infty e^{-x}dxdt \\
&= \int_0^\infty e^{-x} \int_0^x f'(t)dtdx \\
&= \int_0^\infty e^{-x}(f(x) - f(0))dx \\
&= \mathrm{E}[f(Z)] - \mathrm{E}[f(0)] = \mathrm{E}[f(Z)] - f(0)
\end{aligned}$$

For the second statement we use that

$$\begin{aligned}
\int_t^\infty e^{-x}dx &= \int_t^\infty \frac{d}{dx}\left(-e^{-x}\right)dx \\
&= \lim_{s\to\infty} -e^{-x}|_{x=t}^{x=s} = e^{-t}.
\end{aligned}$$

The third equality is obtained by interchanging the order of integration. This proves the first statement.
For the second statement, suppose that W is a non-negative random variable such that $\mathrm{E}\mathcal{C}f(W) = 0$ for all absolutely continuous functions f with $\mathrm{E}|f'(Z)| < \infty$. In particular, the functions $f_k(x) = x^k$ are in this family, and thus

$$\mathrm{E}[f_k'(W)] = \mathrm{E}[f_k(W)] - f_k(0),$$

or equivalently,

$$k\mathrm{E}[W^{k-1}] = \mathrm{E}[W^k] - 0^k = \mathrm{E}[W^k]$$

This equation recursively determines the moments of W, in such a way that its moments must be equal to the moments of a random variable having the exponential distribution with rate 1. By the method of moments, the distribution of W must thus equal the distribution of the exponential distribution with rate 1. This proves the second statement.

∎

Now, we will combine our results to obtain the following corollary.

Corollary 2.16 *Let $E(x)$ denote the distribution function of the exponential distribution. Let $f_x(w)$ denote the unique solution with $f_x(0) = 0$ of the differential equation (2.10). Then, for any random variable W,*

$$|\mathrm{P}(W \leq x) - E(x)| = |\mathrm{E}\mathcal{C}f_x(W)|. \tag{2.11}$$

Equation (2.11) is obviously obtained by inserting the random variable W in Equation (2.10) and then taking the expected value and the absolute value on both sides of the equality sign.

We have now used Stein's method to put bounds on the error when converging to the standard normal distribution, Poisson distribution and the exponential distribution. In the next section we will use Stein's methods to obtain similar results for the negative binomial distribution.

3 Stein's method for the negative binomial distribution

We have now obtained results of Stein's method for the standard normal distribution, Poisson distribution and exponential distribution. However, to apply it to preferential attachments graphs, we need to consider the negative binomial distribution. This section is based on [Brown and Phillips, 1999].
For $r > 0$ and $0 < p \leq 1$, we say that X has the negative binomial distribution $NB(r, p)$ if

$$\mathrm{P}(X = k) = \binom{k+r-1}{k}(1-p)^k p^r \text{ for } k = 0, 1, ...$$

This is a discrete probability distribution that describes the number of failures in a sequence of independent and identically distributed Bernoulli trials with probability of success p before r successes occur. It has mean $r(1-p)/p$.

3.1 The basic results

By now, we have gotten familiar with the outline of Stein's method. Thus, we start by defining a characterizing operator.

Lemma 3.1 *Let $0 < p < 1$, $r \in \mathbb{N}$. Define the functional operator $\mathcal{G}$ by*

$$\mathcal{G}f(x) = (1-p)(r+x)f(x+1) - xg(x)$$

Then, the following statement holds.

1. *If random variable Z has the negative binomial distribution,*

$$\mathrm{E}\mathcal{G}f(Z) = 0$$

for all bounded functions f on $\mathbb{R}_+$.

2. *If for some random variable W,*

$$\mathrm{E}\mathcal{G}f(W) = 0$$

for all bounded functions f on $\mathbb{R}_+$, then W has the negative binomial distribution.

The operator $\mathcal{G}$ is thus referred to as a characterizing operator of the negative binomial distribution.

As expected, our next step is to find the solution of a difference equation, before we can prove Lemma 3.1. This will be done in the following lemma.

Lemma 3.2 *Let $NB_{r,p}$ denote the probability with respect to a negative binomial distribution with parameters r and p and let $A \subseteq \mathbb{N}$. Then, there exists a unique solution f_A of*

$$\mathcal{G}f(k) = \mathbb{I}[k \in A] - NB_{r,p}(A) \tag{3.1}$$

with $f_A(0) = 0$. It is given by

$$f_A(k) = \frac{1}{kNB_{r,p}(\{k\})}\left[NB_{r,p}(A \cap U_k) - NB_{r,p}(A)NB_{r,p}(U_k)\right],$$

where $U_k = 0, 1, ..., k-1$.

Proof. Since the solution is defined recursively, the solution is obviously unique under the initial condition $f_A(0) = 0$.
We insert the supposed solution in the definition of $\mathcal{G}f(k)$ to show that Equation 3.1 holds for this function f_A.

$$\begin{aligned}\mathcal{G}f_A(k) =&(1-p)(r+k)f_A(k+1) - kf_A(k)\\ =&\frac{(1-p)(r+k)}{(k+1)NB_{r,p}(\{k+1\})}\left[NB_{r,p}(A\cap U_{k+1}) - NB_{r,p}(A)NB_{r,p}(U_{k+1})\right]\\ &-\frac{k}{kNB_{r,p}(\{k\})}\left[NB_{r,p}(A\cap U_k) - NB_{r,p}(A)NB_{r,p}(U_k)\right]\\ =&\frac{1}{NB_{r,p}(\{k\})}\left[NBr,p(\{k\})\mathbb{I}[k\in A] - NBr,p(\{k\})NB_{r,p}(A)\right]\\ =&\mathbb{I}[k\in A] - NB_{r,p}(A)\end{aligned}$$

This shows that f_A indeed solves equation 3.1. ∎

Proof. We will start with the first statement of the lemma.
Assume Z is a random variable with the negative binomial distribution. We want to show that

$$\mathrm{E}\mathcal{G}f(Z) = 0,$$

for all bounded functions f, or, equivalently,

$$\mathrm{E}(1-p)(r+Z)g(Z+1) = \mathrm{E}Zg(Z).$$

$$\begin{aligned}\mathrm{E}(1-p)(r+Z)g(Z+1) &= \sum_{k=0}^{\infty}\binom{k+r-1}{k}(1-p)^k p^r(1-p)(r+k)g(k+1)\\ &= \sum_{k=0}^{\infty}\frac{(k+r)!}{(k+1)!(r-1)!}(k+1)(1-p)^{k+1}p^r g(k+1)\\ &= \mathrm{E}Zg(Z)\end{aligned}$$

This proves the statement.
Now, for the second part of the statements, assume W is a random variable such that $\mathrm{E}\mathcal{G}f(W) = 0$ for all bounded functions f. Take $A \subseteq \mathbb{N}$ and take $f_A(k)$ as defined in Lemma 3.2. Then, $\mathrm{E}\mathcal{G}f_A(W) = 0$ implies that

$$0 = \mathrm{E}\mathcal{G}f_A(W) = \mathrm{E}\left[\mathbb{I}[W\in A] - NB_{r,p}(A)\right] = \mathrm{P}(W\in A) - NB_{r,p}(A).$$

Thus,

$$\mathrm{P}(W\in A) = NB_{r,p}(A)$$

for all sets A, and thus W has the negative binomial distribution. ∎

Corollary 3.3 *If W is an integer-valued, positive random variable, then*

$$|\mathrm{P}(W\in A) - NB_{r,p}(A)| = |\mathrm{E}[(1-p)(r+W)f_A(W+1) - Wf_A(W)]|. \tag{3.2}$$

Equation 3.2 is obviously obtained by inserting the random variable W in Equation 3.1 and then taking the expected value and the absolute value on both sides of the equality sign.

3.2 Some necessary definitions

We have now obtained similar results for the negative binomial distribution as we have obtained for the standard normal distribution, Poisson distribution and exponential distribution. However, we need to introduce some definitions that will be used later. We start with the specific mixture distribution, which is approached by W_n.

Definition 3.4 *For $m \in \mathbb{N}$, $\delta > -m$ and U uniform on $[0,1]$, denote the mixture distribution $NB(m+\delta, U^{1/(2+\delta/m)})$ by $K(m,\delta)$.*

This mixture distribution can be interpreted as follows. Let Y be a random variable such that $Y \sim K(m,\delta)$. Then,

$$\mathrm{P}(Y = y) = \int_0^1 \mathrm{P}(NB(m+\delta, x^{1/(2+\delta/m)}) = y)dx.$$

After that, we need to define the *Pólya urn scheme.* This describes an urn with black and white balls. After a ball of a specific colour is drawn, it is returned in the urn and an additional ball of the same colour is added to the urn. We define a random variable $U_{r,n}$ which has the distribution of the number of white balls drawn in $n-1$ draws in a Pólya urn scheme starting with r white balls and 1 black ball. The mass density function of the distribution of $U_{r,n}$ can be determined recursively. Set $U_{r,1}$, or the number of white balls drawn in 0 draws, equal to 0. Then, for $k \geq 1$, given $U_{r,k}$, or the number of white balls drawn so far, the chance of drawing a white ball in the k^{th} draw is given by

$$\mathrm{P}(U_{r,k+1} = U_{r,k} + 1 | U_{r,k}) = \frac{r + U_{r,k}}{r+k} \tag{3.3}$$

Indeed, after $k-1$ draws, there are $r+1+(k-1) = r+k$ balls in the urn, since we started with $r+1$ balls. Since exactly $U_{r,k}$ times a white ball has been added to the urn, there are $r + U_{r,k}$ white balls in the urn. This explains the expression above.
We need to define another notion, namely the *size-biased distribution* with respect to an any integer-valued, non-negative random variable with finite mean X.

Definition 3.5 *Let X be an integer-valued, non-negative random variable with finite mean. We say X^s has the size-biased distribution with respect to X if*

$$\mathrm{P}(X^s = k) = \frac{k\mathrm{P}(X=k)}{\mathrm{E}X}, k = 1, 2, ...$$

Note that for every integer-valued, non-negative random variable with finite mean X, its size-biased distribution is a properly defined distribution, since

$$\begin{aligned}\sum_{k=1}^{\infty} \mathrm{P}(X^s = k) &= \sum_{k=1}^{\infty} \frac{k\mathrm{P}(X=k)}{\mathrm{E}X} \\ &= \frac{\sum_{k=1}^{\infty} k\mathrm{P}(X=k)}{\mathrm{E}X} = \frac{\mathrm{E}X}{\mathrm{E}X} = 1\end{aligned}$$

Furthermore, note that the name ”size-biased distribution” makes sense, because of the factor k in the numerator, which makes probabilities relatively higher for higher values of k.
We need to define one more class of distributions before we can continue.

Definition 3.6 *Let X be a non-negative integer-valued random variable with finite mean and let X^s denote a random variable having the size-biased distribution of X. We say the random variable X^{*r} has the* r-equilibrium transformation *with respect to X if $X^{*r} \stackrel{d}{=} U_{r,X^s}$.*

The distribution of the r-equilibrium transformation thus has the distribution of the number of white balls drawn in a Pólya urn scheme starting with r white balls, where the total number of balls drawn is determined by a random variable having the size-biased distribution of X. Thus, for $k = 0, 1, 2, ...,$

$$\begin{aligned} \mathrm{P}(X^{*r} = k) &= \sum_{l=1}^{\infty} \mathrm{P}(X^s = l) P(U_{r,l} = k) && (3.4) \\ &= \sum_{l=1}^{\infty} \frac{l\mathrm{P}(X = l)}{\mathrm{E}X} P(U_{r,l} = k) && (3.5) \\ &= \sum_{l>k} \frac{l\mathrm{P}(X = l)}{\mathrm{E}X} P(U_{r,l} = k). && (3.6) \end{aligned}$$

The last equality follows from the fact that the probability of drawing k white balls in less than k draws equals 0.

3.3 Taking it further

To illustrate why these distributions were necessary to define, we will make the link between the characterizing operator for the negative binomial distribution and the size-biased distribution. This will allow us to define a second characterizing operator for the negative binomial distribution. For that reason, define the operator $D^{(r)}$ as follows.

$$D^{(r)} f(k) = (k/r + 1) f(k + 1) - (k/r) f(k)$$

Note that

$$\begin{aligned} r(1 - p) D^{(r)} f(k) - pkf(k) &= (1 - p)(k + r) f(k + 1) - k(1 - p) f(k) - pkf(k) && (3.7) \\ &= (1 - p)(k + r) f(k + 1) - kf(k) && (3.8) \\ &= \mathcal{G} f(k). && (3.9) \end{aligned}$$

Lemma 3.7 *If the integer-valued, non-negative random variable X has finite mean $\mu > 0$, X^{*r} has the r-equilibrium distribution of X and g is a function such that the expectations below are well defined, then*

$$\mu \mathrm{E} D^{(r)} g(X^{*r}) = \mathrm{E} X g(X).$$

Proof. We will first show that

$$\mathrm{E} D^{(r)} g(U_{r,n}) = g(n),$$

where $U_{r,n}$ is as defined in Subsection 3.2.
We will prove the statement with induction on n.

1. For $n = 1$, the statement is obvious, since $U_{r,1} = 0$. Furthermore,

   $$D^{(r)} g(0) = (0/r + 1) g(0 + 1) - (0/r) g(0) = g(1),$$

   and thus $\mathrm{E} D^{(r)} g(0) = g(1)$.

2. Now, assume the statement is true for $n = k$, i.e.

$$\mathrm{E}D^{(r)}g(U_{r,k}) = g(k).$$

This can be rewritten as

$$g(n) = \mathrm{E}D^{(r)}g(U_{r,n}) = \mathrm{E}\left[\frac{U_{r,n}+r}{r}g(U_{r,n}+1) - \frac{U_{r,n}}{r}g(U_{r,n})\right] \tag{3.10}$$

$$\Leftrightarrow rg(n) + \mathrm{E}U_{r,n}g(U_{r,n}) = \mathrm{E}(U_{r,n}+r)g(U_{r,n}+1) \tag{3.11}$$

for any function g for which the expectations are well-defined.

3. By Equation (3.3), we have that for a function f,

$$\begin{aligned}
\mathrm{E}f(U_{r,k+1}) &= \mathrm{E}\left[\mathrm{P}(U_{r,k+1} = U_{r,k}+1|U_{r,k})f(U_{r,k}+1) + \mathrm{P}(U_{r,k+1} = U_{r,k}|U_{r,k})f(U_{r,k})\right] \\
&= \mathrm{E}\left[\frac{r+U_{r,k}}{r+k}f(U_{r,k}+1) + \frac{k-U_{r,k}}{r+k}f(U_{r,k})\right] \\
&= \frac{1}{r+k}\left[\mathrm{E}(r+U_{r,k})f(U_{r,k}+1) + (k-U_{r,k})f(U_{r,k})\right]
\end{aligned}$$

We insert our induction hypothesis as stated in Equation (3.10) in this equality, to find

$$\begin{aligned}
\mathrm{E}f(U_{r,k+1}) &= \frac{1}{r+k}\left[rf(k) + \mathrm{E}U_{r,k}f(U_{r,k})\right] + \mathrm{E}\left(1 - \frac{U_{r,k}+r}{r+k}\right)f(U_{r,k}) \\
&= \frac{r}{r+k}f(k) + \frac{1}{r+k}\mathrm{E}U_{r,k}f(U_{r,k}) - \frac{1}{r+k}\mathrm{E}U_{r,k}f(U_{r,k}) + \frac{k}{r+k}\mathrm{E}f(U_{r,k}) \\
&= \frac{r}{r+k}f(k) + \frac{k}{r+k}\mathrm{E}f(U_{r,k}).
\end{aligned}$$

This holds for all f such that the expectations are well-defined, so we take $f = D^{(r)}g$ to obtain

$$\begin{aligned}
\mathrm{E}D^{(r)}g(U_{r,k+1}) =& \frac{r}{r+k}D^{(r)}g(k) + \frac{k}{k+r}\mathrm{E}D^{(r)}g(U_{r,k}) \\
=& \frac{r}{r+k}D^{(r)}g(k) + \frac{k}{k+r}g(k) \\
=& \frac{r}{r+k}\left[\frac{r+k}{r}g(k+1) - \frac{k}{r}g(k)\right] + \frac{k}{k+r}g(k) \\
=& g(k+1),
\end{aligned}$$

which proves the statement.

Thus, we have shown that

$$\mathrm{E}D^{(r)}g(U_{r,n}) = g(n)$$

for all functions g such that the expectation is well-defined. Now, we have that

$$\begin{aligned}
\mu\mathrm{E}D^{(r)}g(X^{*r}) &= \mu\mathrm{E}D^{(r)}g(U_{r,X^s}) \\
&= \mu\sum_{k=1}^{\infty}\mathrm{E}D^{(r)}g(U_{r,k})\mathrm{P}(X^s = k) \\
&= \mu\sum_{k=1}^{\infty}g(k)\mathrm{P}(X^s = k) \\
&= \mu\sum_{k=1}^{\infty}g(k)\frac{k\mathrm{P}(X = k)}{\mathrm{E}X} \\
&= \mathrm{E}Xg(X).
\end{aligned}$$

This proves the lemma. ∎

This lemma allows us to find a new way to uniquely identify a random variable with the negative binomial distribution.

Corollary 3.8 *If the integer-valued random variable $X \geq 0$ is such that $\mathrm{E}X = r(1-p)/p = \mu$ for some $0 < p < 1$ and X^{*r} is a random variable having the r-equilibrium distribution with respect to X, then $X \sim NB(r,p)$ if and only if*

$$X \stackrel{\mathrm{d}}{=} X^{*r}.$$

We will only prove the 'only if'-part of the statement. For the second part, we refer to [Ross, 2013]. This proof makes use of the method of moments.
Proof. Assume $X \stackrel{\mathrm{d}}{=} X^{*r}$. Then, we need to show that

$$\mathrm{E}\mathcal{G}g(X^{*r}) = 0$$

for all bounded functions g. From Equation (3.7) we know that it is equivalent to show that

$$\mathrm{E}\left[r(1-p)D^{(r)}g(X^{*r}) - pX^{*r}g(X^{*r})\right] = 0$$

for all bounded functions g. Take g an arbitrary bounded function.

$$\begin{aligned}\mathrm{E}\left[r(1-p)D^{(r)}g(X^{*r}) - pX^{*r}g(X^{*r})\right] &= p\mu\mathrm{E}D^{(r)}g(X^{*r}) - p\mathrm{E}X^{*r}g(X^{*r}) \\ &= p\mathrm{E}Xg(X) - p\mathrm{E}X^{*r}g(X^{*r}) = 0\end{aligned}$$

In the second equality we used Lemma 3.7. In the last equality we used our assumption that $X \stackrel{\mathrm{d}}{=} X^{*r}$. ∎

This corollary suggests that the distance between a random variable X and a random variable X^{*r} with the r-equilibrium distribution with respect to X, might give us information on how well X approaches a random variable with the negative binomial distribution. Indeed, such a relation exists and it is illustrated by the following theorem, which is proven in [Ross, 2013].

Theorem 3.9 *Let X be a non-negative integer-valued random variable with $\mathrm{E}X = \mu$. Let $r \in \mathbb{N}$ and let X^{*r} be a random variable with the r-equilibrium distribution with respect to X. If $p = r/(r+\mu)$, then*

$$d_{TV}(X, NB(r,p)) \leq 2(e\max\{1,r\}+1)\mathrm{P}(X^{*r} \neq X).$$

Thus, one can find the total variation distance between a random variable X and a random variable having the normal distribution by constructing a random variable with the r-equilibrium distribution with respect to X and then calculating $\mathrm{P}(X^{*r} \neq X)$. This is a very important result, which is at the basis of the results in Section 6.

4 Stein's method for the geometric distribution applied to the uniform attachment random graph model

Before we can apply Stein's method for the negative binomial distribution to preferential attachment graphs, which is the main aim of this article, we will apply Stein's method for the geometric distribution to random graphs. The geometric distribution is a special case of the negative binomial distribution with $r = 1$. Since we have already obtained Stein's method for the negative binomial distribution, it is easy to obtain Stein's method for the geometric distribution.

Remark 4.1 *In this entire section, we use the geometric distribution with support starting at* 0*. In some references, the authors work with the geometric distribution with support starting at* 1*. Because of this, slight differences in obtained results can be observed. However, equivalence of these results is easy to show.*

4.1 Stein's method for the geometric distribution

Recall that in Corollary 3.8 we showed that an integer-valued random variable $X \geq 0$ with $\mathrm{E}X = r(1-p)/p = \mu$ for some $0 < p < 1$, has the negative binomial distribution with parameters r and p if and only if

$$X \sim X^{r*}.$$

Thus, we know that an integer-valued random variable $X \geq 0$ with $\mathrm{E}X = (1-p)/p = \mu$ for some $0 < p < 1$, has the geometric distribution with parameter p if and only if

$$X \sim X^{1*}.$$

However, the expression for X^{1*} can be simplified with the following lemma.

Lemma 4.2 *Let* X *be a non-negative integer-valued random variable with finite mean. Then, random variable* X^e *has the* 1*-equilibrium distribution of* X*, or in short the equilibrium distribution of* X*, if and only if*

$$P(X^e = k) = \frac{P(X > k)}{\mathrm{E}X}.$$

Proof. Let X^s denote the size-biased distribution of X. Recall that X^e has the 1-equilibrium distribution if $X^e \sim U_{1,X^s}$. From Equation (3.4) we know that this is equivalent to stating that for $k = 0, 1, 2, ...$

$$P(X^e = k) = \sum_{l > k} \frac{l\mathrm{P}(X = l)}{\mathrm{E}X} P(U_{1,l} = k)$$

Now, we claim that for all $k \in \{1, ..., l\}$,

$$P(U_{1,l} = k) = \frac{1}{l},$$

or in other words, that $U_{1,l}$, or the number of white balls drawn in $l-1$ draws from a Pólya urn starting with 1 white and 1 black ball, has the uniform distribution on $\{0, 1, ..., l-1\}$. We show this by showing that the number of white balls in the urn after $m-1$ draws is uniform on $\{1, 2, ..., m\}$ by induction on m.

1. After the first draw, there are 3 balls in the urn, of which either 1 or 2 white balls. Both possibilities have probability $\frac{1}{2}$. This shows the statement for $m = 2$.

2. Assume the statement is true for $m = k$, i.e. after $k-1$ draws the number of white balls in the urn is uniform on $\{1, 2, ..., k\}$.

3. Then, we need to prove the statement for $m = k+1$, i.e. we need to show that after k draws the number of white balls in the urn is uniform on $\{1, 2, ..., k+1\}$. Note that drawing r balls is equivalent to drawing $r-1$ balls and then drawing another ball. By the induction statement, for any r such that $1 \leq r \leq k$, the probability of having r white balls in the urn equals $\frac{1}{k}$. The probability of drawing $r-1$ white balls out of k draws

is the probability of drawing $r-2$ white balls out of $k-1$ draws and then drawing 1 white ball, plus the probability of drawing $r-1$ white balls out of $k-1$ draws and then drawing a black ball. The probability of drawing $r-2$ white balls out of $k-1$ draws equals $\frac{1}{k}$ by the induction statement. Similarly, the probability of drawing $r-1$ white balls out of $k-1$ draws equals $\frac{1}{k}$. The probability of drawing a white ball after drawing $r-2$ white balls out of $k-1$ draws equals $\frac{r-1}{k+1}$. The probability of drawing a black ball after drawing $r-1$ white balls out of $k-1$ draws equals $\frac{k+1-r}{k+1}$. Thus, the probability of having r white balls in the urn after k draws, or equivalently, the probability of drawing $r-1$ white balls out of k draws equals

$$\frac{1}{k}\frac{r-1}{k+1} + \frac{1}{k}\frac{k+1-r}{k+1} = \frac{1}{k+1}.$$

Since r was arbitrary, this shows that after k draws the number of white balls in the urn is uniform on $\{1, 2, ..., k+1\}$.

Thus, for all k,

$$P(U_{1,l} = k) = \frac{1}{l}.$$

This means that

$$\begin{aligned} P(X^e = k) &= \sum_{l>k} \frac{l\mathrm{P}(X=l)}{\mathrm{E}X} P(U_{1,l} = k) \\ &= \sum_{l>k} \frac{l\mathrm{P}(X=l)}{l\mathrm{E}X} \\ &= \frac{\mathrm{P}(X>k)}{\mathrm{E}X} \end{aligned}$$

This proves the lemma. ∎

Thus, we can conclude that a random variable X has the geometric distribution if and only if

$$\mathrm{P}(X=k) = P(X^e = k) = \frac{\mathrm{P}(X>k)}{\mathrm{E}X}.$$

This statement can again be rephrased by the following lemma.

Lemma 4.3 *Let X be a non-negative, integer-valued random variable with finite mean. An integer-valued random variable X^e has the equilibrium distribution with respect to X if and only if, for all bounded f and $\Delta f(x) = f(x+1) - f(x)$, we have*

$$\mathrm{E}f(X) - f(0) = \mathrm{E}X\mathrm{E}[\Delta f(X^e)]. \tag{4.1}$$

Proof. Note that

$$\begin{aligned} \mathrm{E}f(X) - f(0) &= \mathrm{E}(f(X) - f(X-1)) + \mathrm{E}(f(X-1) - f(X-2)) + ... + \mathrm{E}(f(2) - f(1)) + \mathrm{E}(f(1) - f(0)) \\ &= \mathrm{E}\left[\sum_{k=0}^{X-1} \Delta f(k)\right] \\ &= \sum_{l=0}^{\infty} \mathrm{P}(X=l) \left(\sum_{k=0}^{l-1} \Delta f(k)\right) \end{aligned}$$

Note that for index k fixed, the term $\Delta f(k)\mathrm{P}(X = l)$ appears for all values of l such that $l > k$. This leads to the following equality.

$$\begin{aligned}
\mathrm{E}f(X) - f(0) &= \sum_{k=0}^{\infty} \Delta f(k)\mathrm{P}(X > k) \\
&= \sum_{k=0}^{\infty} \Delta f(k)\mathrm{E}X\mathrm{P}(X^e = k) \\
&= \mathrm{E}X\mathrm{E}\left[\Delta f(X^e)\right].
\end{aligned}$$

The second equality follows from Lemma 4.2. This proves the statement. ∎

The result above is used to prove the following lemma.

Lemma 4.4 *Let W be a non-negative, integer-valued random variable such that $\mathrm{E}W = (1-p)/p$ for some p such that $0 < p \leq 1$ and let W^e a random variable having the equilibrium distribution with respect to W. Let X be a random variable with the geometric distribution with parameter p. Then,*

$$d_{TV}(W, X) \leq 2(1-p)\mathrm{P}(W \neq W^e).$$

This result is proven in [Peköz et al., 2013]. Note that it is a stronger bound than the bound that follows from Theorem 3.9.

We will need one more result before we can apply our findings to random graphs.

Lemma 4.5 *Let W be an integer-valued random variable, and let W^s have the size-biased distribution of W. Then, if $W \geq 0$ with support at 0, and we define the random variable W^e such that conditional on W^s, W^e has the uniform distribution on the integers $\{0, 1, ..., W^s - 1\}$, then W^e has the equilibrium distribution with respect to W.*

Proof. Note that by the assumptions in the lemma,

$$\mathrm{P}(W^e = k | W^s) = \begin{cases} \frac{1}{W^s} \text{ for } 1 \leq k \leq W^s - 1 \\ 0 \text{ elsewhere} \end{cases}$$

By Lemma 4.3, we need to show that for all bounded f with $\Delta f(x) = f(x+1) - f(x)$, we have

$$\mathrm{E}f(W) - f(0) = \mathrm{E}W\mathrm{E}[\Delta f(W^e)]. \tag{4.2}$$

Remember that W^s has the size-biased distribution with respect to W if for $k = 0, 1, ...$

$$\mathrm{P}(W^s = k) = \frac{k\mathrm{P}(W = k)}{\mathrm{E}W}.$$

Then,

$$\begin{aligned}
\mathrm{E}f(W) - f(0) &= \mathrm{E}\left[\sum_{i=0}^{W-1} \Delta f(i)\right] \\
&= \sum_{k=0}^{\infty} \mathrm{P}(W=k) \sum_{i=0}^{k-1} \Delta f(i) \\
&= \mathrm{E}W \sum_{k=0}^{\infty} \frac{k\mathrm{P}(W=k)}{\mathrm{E}W} \frac{1}{k} \sum_{i=0}^{k-1} \Delta f(i) \\
&= \mathrm{E}W \sum_{k=0}^{\infty} \mathrm{P}(W^s=k) \frac{1}{k} \sum_{i=0}^{k-1} \Delta f(i) \\
&= \mathrm{E}W\mathrm{E}\left[\frac{1}{W^s} \sum_{i=0}^{W^s-1} \Delta f(i)\right] \\
&= \mathrm{E}W\mathrm{E}\left[\sum_{i=0}^{\infty} \mathrm{P}(W^e = i|W^s)\Delta f(i)\right] \\
&= \mathrm{E}W\mathrm{E}[\Delta f(W^e)].
\end{aligned}$$

This proves the statement.

∎

4.2 Application to the uniform attachment random graph model

We will now use the results we obtained applying Stein's method to the geometric distribution to show properties of the uniform attachment random graph model.
We will firstly explain the uniform attachment random graph model. Let G_n be a directed random graph on n nodes defined by the following recursive construction. Define G_1 as one node with a single loop. One end of the loop corresponds to the out-degree of the node and the other end corresponds to the in-degree of the node. For $m > 1$, given G_{m-1}, add node m along with an edge directed from m to a node chosen uniformly at random along the m nodes present. This can also be node m itself. This is among the easiest types of random graphs that can be constructed.
In [Chen et al., 2010], it has been shown that the random variable W which equals the in-degree of a node chosen uniformly at random from G_n converges in distribution to a geometric distribution with parameter $1/2$ as $n \to \infty$. In the following theorem, we will obtain a bound on the total variation distance between the distribution of W and the distribution of a random variable with the geometric distribution.

Theorem 4.6 *Let W be the in-degree of a node chosen uniformly at random from the random graph G_n, generated according to uniform attachment. Let X be a random variable with the geometric distribution with parameter $1/2$. Then,*

$$d_{TV}(W, X) \leq \frac{1}{n}$$

Proof. Let X_i have a Bernoulli distribution, independent of all else, with parameter

$$\mu_i = \frac{1}{n-i+1}$$

and let N be a random variable that is uniform on the integers $\{1, 2, ..., n\}$. Note that then also $M = n + 1 - N$ is uniform on the integers $\{1, 2, ..., n\}$ and imagine M is the random variable describing the node that is selected. Then, all the nodes added after M and M itself (this is a total of N nodes) can contribute to the in-degree of node. Node n has a chance of $\mu_1 = 1/n$ of connecting to node M, node $n - 1$ has a chance of $\mu_2 = 1/(n - 1)$ of connecting to node M, etc. Thus, we can conclude

$$W = \sum_{i=1}^{N} X_i.$$

We will now show that

$$\sum_{i=1}^{N-1} X_i$$

has the equilibrium distribution with respect to W.
Using Lemma 4.3, we know that this is equivalent to showing that for all bounded f and $\Delta f(x) = f(x + 1) - f(x)$ we have

$$\mathrm{E}f(X) - f(0) = \mathrm{E}X\mathrm{E}[\Delta f(X^e)]. \tag{4.3}$$

For this, we will use the following relation.

$$\begin{aligned}\mathrm{E}f\left(\sum_{i=1}^{m} X_i\right) &= \mu_m \mathrm{E}f\left(\sum_{i=1}^{m-1} X_i + 1\right) + (1 - \mu_m)\mathrm{E}f\left(\sum_{i=1}^{m-1} X_i\right) \\ &= \mu_m \mathrm{E}\Delta f\left(\sum_{i=1}^{m-1} X_i\right) + \mathrm{E}f\left(\sum_{i=1}^{m-1} X_i\right)\end{aligned}$$

This yields

$$\mu_m \mathrm{E}\Delta f\left(\sum_{i=1}^{m-1} X_i\right) = \mathrm{E}\left[f\left(\sum_{i=1}^{m} X_i\right) - f\left(\sum_{i=1}^{m-1} X_i\right)\right]. \tag{4.4}$$

Furthermore, note that for every bounded function g with $g(0) = 0$, the following holds.

$$\mathrm{E}\left[\frac{\Delta g(N-1)}{\mu_N}\right] = \sum_{i=1}^{n} \frac{1}{n}(n - i + 1)(g(i) - g(i - 1)) \tag{4.5}$$

$$= \frac{1}{n}\left[-g(0) + \sum_{i=1}^{n-1}((n - i + 1) - (n - i))g(i) + g(n)\right] \tag{4.6}$$

$$= \frac{1}{n}\sum_{i=1}^{n} g(n) = \mathrm{E}g(N) \tag{4.7}$$

Then, using Equation (4.4) and assuming $f(0) = 0$, we find that

$$\begin{aligned}
\mathrm{E}\left[\Delta f\left(\sum_{i=1}^{N-1} X_i\right)\right] &= \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}\left[\Delta f\left(\sum_{i=1}^{k-1} X_i\right)\right] \\
&= \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}\left[\frac{1}{\mu_k}\left[f\left(\sum_{i=1}^{k} X_i\right) - f\left(\sum_{i=1}^{k-1} X_i\right)\right]\right] \\
&= \mathrm{E}\left[\frac{1}{\mu_N}\left[f\left(\sum_{i=1}^{N} X_i\right) - f\left(\sum_{i=1}^{N-1} X_i\right)\right]\right] \\
&= \sum_{x_1=0}^{1}\cdots\sum_{x_n=0}^{1}\mathrm{P}(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)\mathrm{E}\left[\frac{1}{\mu_N}\left[f\left(\sum_{i=1}^{N} x_i\right) - f\left(\sum_{i=1}^{N-1} x_i\right)\right]\right] \\
&= \sum_{x_1=0}^{1}\cdots\sum_{x_n=0}^{1}\mathrm{P}(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)\mathrm{E}\left[f\left(\sum_{i=1}^{N} x_i\right)\right] \\
&= \mathrm{E}\left[f\left(\sum_{i=1}^{N} X_i\right)\right] \\
&= \mathrm{E}f(W)
\end{aligned}$$

For the fifth equality we use Equation (4.5) with

$$g(k) = f\left(\sum_{i=1}^{k} x_i\right).$$

Note that since $f(0) = 0$, $g(0) = 0$.
Furthermore, note that

$$\begin{aligned}
\mathrm{E}W &= \mathrm{E}\left[\sum_{i=1}^{N} X_i\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{n} X_i \mathbb{I}[i \leq N]\right] \\
&= \sum_{i=1}^{n}\mathrm{E}[X_i]\mathrm{P}(i \leq N) \\
&= \sum_{i=1}^{n}\frac{1}{n-i+1}\cdot\frac{n-i+1}{n} \\
&= \sum_{i=1}^{n}\frac{1}{n} \\
&= 1.
\end{aligned}$$

Thus,

$$\mathrm{E}f(W) - f(0) = 1\cdot\mathrm{E}f(W) = \mathrm{E}W\mathrm{E}\Delta f\left(\sum_{i=1}^{N-1} X_i\right) \tag{4.8}$$

for all bounded f such that $f(0) = 0$. However, in order to use Lemma 4.3, we should show that Equation (4.3) holds for all bounded f. Therefore, assume f is such a function. Then,

$g(x) := f(x) - f(0)$ is a function such that $g(0) = 0$, so Equation (4.8) holds for g. Thus, noting that $\Delta h(x) + C = \Delta h(x)$ for all functions h and constants C, we have that

$$\begin{aligned}
\mathrm{E}f(W) - f(0) =& \mathrm{E}[f(W) - f(0)] = \mathrm{E}g(W) \\
=& \mathrm{E}W\mathrm{E}\Delta g\left(\sum_{i=1}^{N-1} X_i\right) \\
=& \mathrm{E}W\mathrm{E}\Delta\left[g\left(\sum_{i=1}^{N-1} X_i\right) + f(0)\right] = \mathrm{E}W\mathrm{E}\Delta f\left(\sum_{i=1}^{N-1} X_i\right).
\end{aligned}$$

This shows that for all bounded f, Equation (4.3) holds.
This implies, using Lemma 4.3, that

$$\sum_{i=1}^{N-1} X_i$$

has the equilibrium distribution with respect to W, so that

$$\mathrm{P}[W \neq W^e] \leq \mathrm{P}[N = n] = \frac{1}{n}.$$

Using Lemma 4.4, we find that

$$d_{TV}(W, X) \leq 2(1/2)\mathrm{P}(W \neq W^e) = \frac{1}{n}.$$

This is the desired result. ∎

We have thus quantified how well the geometric distribution with parameter $1/2$ is approximated by the random variable describing the in-degree of a node chosen uniformly at random from a random graph with n nodes, generated according to uniform attachment. We can conclude that the total variation distance between these distributions is at most $1/n$. In the next section we will quantify the error in the approximation of another type of random graph.

5 Stein's method for the geometric distribution applied to preferential attachment graphs

In the real world, many dynamical networks are found. These are networks that can be described by graphs that are dependent on time. Examples of these real-world dynamical networks are the structure of the Internet, social networks and biological networks. A striking feature that many of these networks possess, is that the degree of the vertices is distributed according to a power law sequence. This means that if the number of vertices in the graph describing the network is high, the proportion of vertices having degree k approximately decays as $c_\gamma k^{-\gamma}$, for some constant c_γ and $\gamma > 1$. Preferential attachment graphs are a specific type of random graphs that do quite a good job resembling real-life dynamical networks with a power-law degree distribution. Thus, they can help us to give insight in explaining the dynamics in the networks we see around us. The definition of preferential attachment graphs is based on [Van Der Hofstad, 2009]. The results concerning Stein's method for the geometric distribution are based on [Peköz et al., 2013].

Preferential attachment models are models in which sequentially vertices are added to a graph. They are then connected to a number of older vertices. These edges are attached to another vertex with a probability proportional to the degree of the receiving vertex at that time. This means that the vertices with a higher degree are favoured in the process, and thus these models are sometimes referred to as 'rich-get-richer' models. It is intuitive that preferential attachment is also seen in real-life situations. For example, a newcomer in a social network is more likely to interact with social people that already know many people than with the more isolated people.

5.1 Definition of the preferential attachment model

The preferential attachment model is a stochastic process that produces a graph sequence. We denote this sequence by $\left(G_k^{m,\delta}\right)_{k\geq 1}$, which for every iterate k yields a graph of k vertices and mk edges with $m \in \mathbb{N}$ and $\delta > -m$

We will start by defining the model for $m = 1$. We denote the vertices of $G_k^{1,\delta}$ by $\{v_1^1, ..., v_k^1\}$. Furthermore, denote the degree of vertex v_i^1 in $G_k^{1,\delta}$ by $W_{k,i}^{\delta}$. A self-loop increases the degree by 2. Define $G_1^{1,\delta}$ as vertex v_1^1 with one edge connecting v_1^1 to itself.

$G_{k+1}^{1,\delta}$ is obtained from $G_k^{1,\delta}$ with use of a growth rule. We add a single vertex v_k^1 with a single edge to the graph. This edge is connected to a second end point from the set $\{v_1^1, ..., v_k^1, v_{k+1}^1\}$, with probability connecting to a specific vertex proportional to the total degree of that vertex plus δ. Note that the smaller δ is, the stronger is the favoring effect of the vertices with high degree. Initially, vertex v_{k+1}^1 has degree 1. Thus, if we define $X_{i,k+1}$ as the indicator variable of the event that v_{k+1}^1 connects to v_i^1 with $i \in \{1, ..., k+1\}$,

$$\mathrm{P}(X_{i,k+1} = 1 | G_k^{1,\delta}) = \begin{cases} \frac{1+\delta}{k(2+\delta)+(1+\delta)} & \text{for } i = k+1 \\ \frac{W_{k,i}^{\delta}+\delta}{k(2+\delta)+(1+\delta)} & \text{for } i < k+1 \end{cases} . \tag{5.1}$$

The numerator of the fractions is explained by the defined weight for the different vertices. The denominator has this value, since in total $2k$ endpoints of edges are present in the first k vertices (k edges have been added), every of the k vertices gets δ added to its weight, and the $(k+1)^{\text{th}}$ vertex has weight $1+\delta$. Thus, to make sure the sum of all probabilities equals 1, the denominator should have this value.

Now that we have defined the stochastic process to generate a sequence $\left(G_k^{1,\delta}\right)_{k\geq 1}$. The graph $G_k^{1,\delta}$ will be used to define the graph $G_k^{m,\delta}$ with $m \in \mathbb{N}$.

Start with $G_{mk}^{1,\delta/m}$ with vertices $\{v_1^1, ..., v_{mk}^1\}$. Now collapse the vertices $\{v_1^1, ..., v_m^1\}$ in $G_{mk}^{1,\delta/m}$ to become the vertex v_1^m in $G_k^{m,\delta}$. In this process, let all edges that are connected to any of the vertices $\{v_1^1, ..., v_m^1\}$ in $G_{mk}^{1,\delta/m}$ be connected to the new vertex v_1^m in $G_k^{m,\delta}$. Similarly, collapse vertices $\{v_{m+1}^1, ..., v_{2m}^1\}$ in $G_{mk}^{1,\delta/m}$ to become the vertex v_2^m in $G_k^{m,\delta}$. In general, we collapse vertices $\{v_{(j-1)m+1}^1, ..., v_{(j)m}^1\}$ in $G_{mk}^{1,\delta/m}$ to become the vertex v_j^m in $G_k^{m,\delta}$. This defines the construction of $G_k^{m,\delta}$. The resulting graph obviously has mk edges and k vertices. Note that in this way, vertices are able to connect to themselves more than once. In fact, the first vertex consists of m loops and has in- and out-degree at least m.

Furthermore, note that an edge in $G_k^{m,\delta}$ is attached to vertex v_i^1 with probability proportional to its degree plus δ/m. Since vertex v_j^m in $G_k^{m,\delta}$ is formed by collapsing vertices $\{v_{(j-1)m+1}^1, ..., v_{(j)m}^1\}$ in $G_{mk}^{1,\delta/m}$, an edge in $G_k^{m,\delta}$ is attached to v_j^m with probability proportional to the total weight of the m vertices $\{v_{(j-1)m+1}^1, ..., v_{(j)m}^1\}$. The sum of the degrees of vertices $\{v_{(j-1)m+1}^1, ..., v_{(j)m}^1\}$ is equal to the degree of v_j^m, and thus this probability is proportional to the degree of v_j^m plus δ.

5.2 Applying Stein's method for the geometric distribution

In the rest of this section we will only consider $G_n^{1,0}$. We will call this graph G_n and call its vertices v_i for $i = 1, 2, ..., n$. Let random variable W_n equal the in-degree of a node chosen uniformly at random from G_n. The distribution of this random variable converges to a specific distribution, called the *Yule-Simon distribution.*

Definition 5.1 *A random variable Z has the Yule-Simon distribution if for $k = 0, 1, 2, ...$*

$$\mathrm{P}(Z = k) = \frac{4}{(k+1)(k+2)(k+3)}.$$

Lemma 5.2 *The Yule-Simon distribution is a properly defined distribution. In other words, if Z is a random variable having the Yule-Simon distribution,*

$$\sum_{k=0}^{\infty} \mathrm{P}(Z = k) = 1.$$

Proof. Note that

$$\frac{1}{(k+1)(k+2)(k+3)} = \frac{1}{2}\left(\frac{k!2!}{(k+3)!}\right).$$

Furthermore, by the properties of the Beta Function, we know that

$$\frac{k!2!}{(k+3)!} = \int_0^1 x^k(1-x)^2 dx.$$

Then, we observe that

$$\begin{aligned}
\sum_{k=0}^{\infty} \mathrm{P}(Z = k) &= \sum_{k=0}^{\infty} \frac{4}{(k+1)(k+2)(k+3)} \\
&= 2\sum_{k=0}^{\infty} \int_0^1 x^k(1-x)^2 dx \\
&= 2\int_0^1 \left(\sum_{k=0}^{\infty} x^k\right)(1-x)^2 dx \\
&= 2\int_0^1 (1-x) dx = 1,
\end{aligned}$$

where the fourth equality follows from convergence of the geometric series. This proves our statement. ∎

A property of the Yule-Simon distribution that relates preferential attachment graphs to the geometric distribution, is the following. It is proven in [Peköz et al., 2013].

Lemma 5.3 *Let U be a random variable with a uniform distribution on $(0, 1)$. Let Z be the random variable with the mixture distribution consisting of the geometric distribution with parameter distributed as $\sqrt{U}$, or symbolically,*

$$Z \sim Ge(\sqrt{U}).$$

Then, Z has the Yule-Simon distribution.

In this section we will focus on proving the following theorem.

Theorem 5.4 *Let $W_{n,i}$ be the total degree of vertex i in the preferential attachment graph consisting of n vertices and let I be uniform on $\{1, 2, ..., n\}$ independent of $W_{n,i}$. If Z has the Yule-Simon distribution, then*

$$d_{TV}(W_{n,I} - 1, Z) \leq \frac{C \log n}{n}$$

for some constant C independent of n.

The following results will be used in the proof of this theorem.

Lemma 5.5 *Let $W_{n,i}$ be the total degree of vertex i in the preferential attachment graph consisting of n vertices. Then,*

$$\left| \mathrm{E}W_{n,i} - \sqrt{\frac{n}{i}} \right| \leq C \frac{\sqrt{n}}{i^3} \text{ and } \left| \frac{1}{\mathrm{E}W_{n,i}} - \sqrt{in} \right| \leq \frac{C}{\sqrt{ni}}$$

with C a constant independent of n and i.

This lemma is proven in [Bollobás et al., 2001].

Lemma 5.6 *Let W and V be random variables and let X be a random element defined on the same probability space. Then,*

$$d_{TV}(W, V) \leq \mathrm{E}d_{TV}(W|X, V|X)$$

Proof. If $f : \mathbb{R} \to [0, 1]$, then

$$|\mathrm{E}[f(W) - f(V)]| = |\mathrm{E}[\mathrm{E}[f(W) - f(V)|X]]| \leq \mathrm{E}|\mathrm{E}[f(W) - f(V)|X]| \leq \mathrm{E}d_{TV}(W|X, V|X).$$

The first inequality follows from Jensen's inequality and the second inequality follows from the definition of $d_{TV}(W|X, V|X)$.

∎

We will now prove another important intermediate result, before we are ready to prove Theorem 5.4.

Theorem 5.7 *Let $W_{n,i}$ be the total degree of vertex i in the preferential attachment graph consisting of n vertices. Then,*

$$d_{TV}(W_{n,i} - 1, Ge(1/\mathrm{E}(W_{n,i}))) \leq \frac{C}{i}$$

with C a constant independent of n and i.

Proof. In the proof of this theorem, we want to use Lemma 4.5, which uses the equilibrium distribution for its results on geometric approximation. The equilibrium distribution with respect to a random variable makes use of the size-biased distribution with respect to that random variable. Therefore, we want to construct a variable having the size-biased distribution of $W_{n,i} - 1$, which is the out-degree of v_i in G_n.
First, for $j \geq i$ define $X_{j,i}$ as the indicator variable of the event that vertex v_j has an outgoing edge connected to vertex v_i in G_j. Thus, we can denote

$$W_{j,i} = 1 + \sum_{k=i}^{j} X_{k,i}.$$

The term 1 covers the out-degree of the vertex, and the other terms cover the in-degree of the vertex. The sum starts at $k = i$, since vertices with an index less than i can not connect to the at that point not yet existing vertex v_i.
Then we have that for $1 \leq i \leq n$,

$$P(X_{i,i} = 1|G_{i-1}) = \frac{1}{2i-1} \tag{5.2}$$

and for $1 \leq i < j \leq n$,

$$P(X_{j,i} = 1|G_{j-1}) = \frac{W_{j-1,i}}{2j-1}. \tag{5.3}$$

This follows from Equation 5.1. To continue our proof, we need the following result.

Proposition 5.8 *Let $X_1, ..., X_n$ be random variables which can take 0 and 1 as value with $P(X_m = 1) = p_m$. For each $m = 1, ..., n$, let $\left(X_j^{(m)}\right)_{j\neq m}$ have the distribution of $(X_j)_{j\neq m}$ conditional on $X_m = 1$. If $X = \sum_{m=1}^{n} X_m$, $\mu = EX$ and K is chosen independently of the other variables with $P(K = k) = p_k/\mu$, then $X^s = \sum_{j\neq K} X_j^{(K)} + 1$ has the size-biased distribution of X.*

Proof.

$$\begin{aligned}
P(X^s = x) &= \sum_{k=1}^{n} P(K = k) \cdot P\left(\sum_{i\neq k} X_i = x - 1 | X_k = 1\right) \\
&= \frac{1}{\mu}\sum_{k=1}^{n} P(X_k = 1) \cdot P\left(\sum_{i\neq k} X_i = x - 1 | X_k = 1\right) \\
&= \frac{1}{\mu}\sum_{k=1}^{n} P\left(\sum_{i\neq k} X_i = x - 1 \bigcap X_k = 1\right) \\
&= \frac{1}{\mu}\sum_{k=1}^{n} \sum_{\substack{A\subseteq\{1,...,n\}\setminus\{k\} \\ \#A=x-1}} P(X_1 = 1 \text{ for } i \in A \cup \{k\}, X_i = 0 \text{ for } i \notin A \cup \{k\}) \\
&= \frac{1}{\mu} \sum_{\substack{A\subseteq\{1,...,n\} \\ \#A=x}} \sum_{k\in A} P(X_1 = 1 \text{ for } i \in A, X_i = 0 \text{ for } i \notin A) \\
&= \frac{x}{\mu} \sum_{\substack{A\subseteq\{1,...,n\} \\ \#A=x}} P(X_1 = 1 \text{ for } i \in A, X_i = 0 \text{ for } i \notin A) \\
&= \frac{x}{\mu} P\left(\sum_{i=1}^{n} X_i = x\right)
\end{aligned}$$

The first equality follows from the definition of X^s. The second equality follows from the distribution of K. The third equality follows from the properties of conditional probabilities. The fifth equality follows from 'including' k in the sets A. The sixth equality follows from the fact that the summed terms do not depend on k and from $\#A = x$. Comparing the last line to the definition of the size-biased distribution proves the statement. ∎

We will now explain how this proposition relates to our problem.

Let's construct step by step how we can use this result for our purpose. This proposition constructs the size-biased distribution of $X = \sum_{m=1}^{n} X_m$. We are interested in the size bias distribution of $W_{n,i} - 1 = \sum_{m=i}^{n} X_{m,i}$. Therefore, we choose $X_m := X_{m,i}$. Then, for $m = 1, ..., n$, p_m is given by $\mathrm{P}(X_m = 1)$, or in our previous notation by $\mathrm{P}(X_{m,i} = 1)$. This is the probability that v_m has an outgoing edge to v_i. It is obvious that $p_m = 0$ for $m < i$.
Furthermore $\left(X_j^{(m)}\right)_{j \neq m}$ has the distribution of $(X_j)_{j \neq m}$ conditional on $X_m = 1$. Or, in our previous notation, $\left(X_j^{(m)}\right)_{j \neq m}$ has the distribution of $(X_{j,i})_{j \neq m}$ conditional on $X_{m,i} = 1$. Or, rephrasing the statement once more, $\left(X_j^{(m)}\right)_{j \neq m}$ has the distribution of the indicator variable of the event that v_j has an outgoing edge to v_i in G_j (for j unequal to m) conditional on the event that v_m has an outgoing edge to v_i.
After that, we will take a look at the meaning of random variable K. K is a random variable with the property that $\mathrm{P}(K = k)$ is proportional to p_k. Thus, for $k < i$ and $k > n$, $\mathrm{P}(K = k) = 0$ and for $i \leq k \leq n$, $\mathrm{P}(K = k)$ is proportional to the probability that v_k has an outgoing edge to v_i. $(*)$ To be precise, $\mathrm{P}(K = k) = \frac{p_k}{\mathrm{E}X} = \frac{\mathrm{E}X_{k,i}}{\mathrm{E}W_{n,i}-1}$.
So, how do we obtain the size-biased distribution of $W_{n,i} - 1$, now that we have we tailored the result of Proposition 5.8? The proposition states that a random variable having the size-biased distribution of $W_{n,i} - 1 = \sum_{m=i}^{n} X_{m,i}$ is given by

$$X_i^s := X^s = \sum_{j \neq K} X_j^{(K)} + 1.$$

Thus, first we sample K with $\mathrm{P}(K = k)$ as stated in $(*)$. Assume $K = k$. After that, we introduce the condition that $X_k = X_{k,i} = 1$. In other words, we add an outgoing edge from v_k to v_i. Then, we sample $X_j^{(k)}$ for $j \neq k$. I.e. we sample the indicator variable $X_{j,i}$ with the condition that $X_{k,i} = 1$. Then, we count the number of vertices connecting to v_i. This equals sum of the sampled values of $X_{j,i}$ with $j \neq k$ plus 1, since $X_{k,i} = 1$.
This procedure has created a random variable with the size-biased distribution of $W_{n,i} - 1$.
To continue, we need better insight in the probabilities mentioned in the construction of X_i^s.

Lemma 5.9 *For $i \leq j$, let $X_{j,i}$ as the indicator variable of the event that vertex v_j has an outgoing edge connected to vertex v_i in G_j and let $W_{j,i}$ be the total degree of vertex i in the preferential attachment graph consisting of j vertices. Then, for $k > j$, we have*

$$\mathrm{P}(X_{j,i} = 1 | X_{k,i} = 1, G_{j-1}) = \frac{1 + W_{j-1,i}}{2j}, \tag{5.4}$$

where we define $W_{i-1,i} = 1$.

Note that according to the earlier definition $W_{i-1,i}$ would be the total degree of vertex i in the preferential attachment graph consisting of $i - 1$ vertices, which makes no sense. Furthermore, observe that it's striking about this statement that an indicator of a connection in an earlier version of the graph ($X_{j,i}$) is conditional on an indicator of a connection in a later version of the graph ($X_{k,i}$).
Observe that the expression for $\mathrm{P}(X_{j,i} = 1 | X_{k,i} = 1, G_{j-1})$ is very similar to the expression of $\mathrm{P}(X_{j,i} = 1 | G_{j-1})$ given earlier. The only difference is that both in the numerator and the denominator, the extra edge from v_k to v_i is taken into consideration. That is, in the numerator, the degree of v_i is increased by 1, since it has one extra incoming edge. In the

denominator the total degree is increased by 1, since there is one extra connection among the vertices v_1 to v_j.
We will now prove Lemma 5.9.
Proof.

We will prove the lemma by Bayes's rule, which states that for events A and B

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

In this rule, consider $X_{j,i} = 1$ as event A and $X_{k,i} = 1$ as event B. $G_{j,i}$ is an extra condition that will not be worked with while using Bayes's rule.
Then we have that

$$P(X_{j,i} = 1|X_{k,i} = 1, G_{j-1}) = \frac{P(X_{j,i} = 1|G_{j-1})P(X_{k,i} = 1|X_{j,i} = 1, G_{j-1})}{P(X_{k,i} = 1|G_{j-1})} \tag{5.5}$$

We will calculate the three probabilities appearing in the right-hand side of Equation (5.5).
First, for $i \leq j$, we have

$$P(X_{j,i} = 1|G_{j-1}) = \frac{W_{j-1,i}}{2j-1},$$

which is a combination of Equation (5.2) and Equation (5.3) with our new definition of $W_{i-1,i}$.
This implies that

$$P(X_{k,i} = 1|G_{j-1}) = E\left(\frac{W_{k-1,i}}{2k-1}|G_{j-1}\right) = \frac{E[W_{k-1,i}|G_{j-1}]}{2k-1}.$$

Similarly,

$$P(X_{k,i} = 1|X_{j,i} = 1, G_{j-1}) = E\left(\frac{W_{k-1,i}}{2k-1}|X_{j,i} = 1, G_{j-1}\right) = \frac{E[W_{k-1,i}|X_{j,i} = 1, G_{j-1}]}{2k-1}.$$

Furthermore, note that by the method of constructing the graph,

$$E(W_{k,i}|G_{k-1}) = W_{k-1,i} + \frac{W_{k-1,i}}{2k-1} = \left(\frac{2k}{2k-1}\right)W_{k-1,i}.$$

This is the sum of the degree of v_i after $k-1$ steps plus the probability that the k^{th} vertex is connected to v_i.
Repeating this reasoning, we find that

$$E(W_{k,i}|G_{k-2}) = \left(\frac{2(k-1)}{2(k-1)-1}\right)\left(\frac{2k}{2k-1}\right)W_{k-2,i}.$$

Thus, for $i, s < k$,

$$E(W_{k,i}|G_{k-s}) = \prod_{m=1}^{s}\left(\frac{2(k-m+1)}{2(k-m+1)-1}\right)W_{k-s,i}.$$

Changing the indices by setting $j-1 = k-s$ and replacing $k-1$ by k, we get

$$E(W_{k-1,i}|G_{j-1}) = \prod_{m=1}^{k-j}\left(\frac{2(k-m)}{2(k-m)-1}\right)W_{j-1,i}.$$

From this we get that

$$E(W_{k-1,i}|G_j) = \prod_{m=1}^{k-j-1}\left(\frac{2(k-m)}{2(k-m)-1}\right)W_{j,i}.$$

However, the only information we use from G_j is $W_{j,i}$. If we know G_{j-1} and we know $X_{j,i} = 1$, we know $W_{j,i} = W_{j-1,i} + 1$. Thus,

$$E(W_{k-1,i}|X_{j,i} = 1, G_{j-1}) = \prod_{m=1}^{k-j-1} \left(\frac{2(k-m)}{2(k-m)-1} \right) (W_{j-1,i} + 1).$$

We now have all the ingredients of the right-hand side of Equation (5.5), so we can insert the values of the conditional probabilities.

$$\begin{aligned} P(X_{j,i} = 1|X_{k,i} = 1, G_{j-1}) &= \frac{P(X_{j,i} = 1|G_{j-1})P(X_{k,i} = 1|X_{j,i} = 1, G_{j-1})}{P(X_{k,i} = 1|G_{j-1})} \\ &= \frac{\frac{W_{j-1,i}}{2j-1} \frac{E[W_{k-1,i}|X_{j,i}=1,G_{j-1}]}{2k-1}}{\frac{E[W_{k-1,i}|G_{j-1}]}{2k-1}} \\ &= \frac{\frac{W_{j-1,i}}{2j-1} \frac{\prod_{m=1}^{k-j-1}\left(\frac{2(k-m)}{2(k-m)-1}\right)(W_{j-1,i}+1)}{2k-1}}{\frac{\prod_{m=1}^{k-j}\left(\frac{2(k-m)}{2(k-m)-1}\right)W_{j-1,i}}{2k-1}} \\ &= \frac{1 + W_{j-1,i}}{2j} \end{aligned}$$

This proves the statement. ∎

With use of this lemma, we have a method to construct $(W_{n,i}|X_{k,i} = 1)$ for any $1 \leq i \leq k \leq n$. Quantities related to this construction are given the superscript k.
First, we generate G_{i-1}^k according to our 'normal' preferential attachment model. The edge from v_k to v_i has nothing to do with this construction yet, since vertices cannot connect to a vertex with a higher index. Thus, for all $m < i$, $X_{m,i} = 0$, independent of the fact that $X_{k,i} = 1$.
Then, if $i \neq k$, vertices v_i and v_k are added to the graph, along with a vertex labeled $v_{i'}$ with an edge to it coming from v_k. Given G_{i-1}^k and these additional vertices and edges, we generate G_i^k by connecting v_i to a vertex chosen from $\{v_1, ..., v_i, v_{i'}\}$ proportional to their degree. Note that in this process, v_i has degree 1 from its out-edge and $v_{i'}$ has degree 1 from the in-edge coming from v_k.
If $i = k$, we only add vertex v_i and $v_{i'}$ and connect v_i to $v_{i'}$. We can not add another edge originating from v_i, since every vertex has out-degree 1. We denote the resulting graph by G_i^i.
For $i < j < k$, we generate the graphs G_j^k recursively from G_{j-1}^k by connecting v_j to a vertex v_l, randomly chosen from the vertices $v_1, ..., v_j, v_{i'}$ proportional to their degree, where j has degree 1 from its out-edge, as always. Of course, v_k is not included in this process, since for $m < k$, $X_{m,k} = 0$ by definition.
Once we have constructed G_{k-1}^k, we note that since v_k already has out-degree 1, we can not connect it to another vertex, so we have to define $G_k^k = G_{k-1}^k$. Now everything is almost back to normal, and for $j = k+1, ..., n$ we generate G_j^k from G_{j-1}^k recursively according to preferential attachment among the vertices $v_1, ..., v_j, v_{i'}$.
The following lemma gives insight in the properties of this construction.

Lemma 5.10 *Let $1 \leq i \leq k \leq n$ and retain the notation and definitions above. Then, the following statements hold.*

1. $\mathcal{L}(W_{n,i}^k + W_{n,i'}^k) = \mathcal{L}(W_{n,i}|X_k = 1)$
2. *For fixed i, if K is a random variable such that for $k \geq i$,*

$$P(K = k) = \frac{EX_{k,i}}{EW_{n,i} - 1},$$

then $W^K_{n,i} + W^K_{n,i'} - 1$ *has the size-biased distribution of* $W_{n,i} - 1$.

3. *Conditional on the event* $\{W^k_{n,i} + W^k_{n,i'} = m+1\}$, *the variable* $W^k_{n,i}$ *is uniformly distributed on the integers* $\{1, ..., m\}$.

4. $W^K_{n,i} - 1$ *has the equilibrium distribution of* $W_{n,i} - 1$.

Proof.

1. This statement follows from the construction of G^k_n combined with Lemma 5.9.

2. This statement follows from statement 1 combined with Proposition 5.8.

3. One can view $W^k_{n,i}$ and $W^k_{n,i'}$ as the number of balls of two colours in a Pólya urn model starting with one ball of each colour. In the proof of Lemma 4.2 we have shown that $W^k_{n,i}$, or the number of white balls after $m-1$ draws is uniform on $1, ..., m$.

4. This statement follows from Lemma 4.5 after shifting the support of W in the assumption from non-negative to positive integers and altering the results accordingly.

∎

∎

We are now going to prove Theorem 5.7.

Proof. We have to show that for $1 \le i \le n$ and with $W_{n,i}$ being the random variable that equals the total degree of vertex v_i in the preferential attachment graph on n vertices,

$$d_{TV}(W_{n,i} - 1, Ge(1/\mathrm{E}(W_{n,i}))) \le \frac{C}{i}$$

with C a constant independent of n and i. For this, we will apply Lemma 4.4 to $W_{n,i} - 1$, or the random variable that equals the in-degree of vertex v_i in the preferential attachment graph on n vertices. This means that we need to find $\mathrm{P}((W_{n,i} - 1) \neq (W_{n,i} - 1)^e)$ for random variable $(W_{n,i} - 1)^e$ that has the equilibrium distribution with respect to $W_{n,i} - 1$.
To achieve this, we will construct two sequences of random variables for each fixed $k = i, ..., n$. The first sequence we will construct will be called $(X^k_{j,i})_{j \ge i}$ and the random variables in this sequence will be distributed as the indicator of the event that vertex v_j connects to vertex v_i in G^k_n. Recall that G^k_n is the graph of n vertices that is constructed according to preferential attachment, conditional on $X^k = 1$. The second sequence will be called $(\tilde{X}^k_{j,i})_{j \ge i}$ and the random variables in this sequence will be distributed as the indicator of the event that vertex v_j connects to vertex v_i in G_n. However, since we will couple these sequences, the index k is still present.
We will use the notation

$$W^k_{j,i} = \sum_{m=i}^{j} X^k_{j,i} \text{ and } \tilde{W}^k_{j,i} = \sum_{m=i}^{j} \tilde{X}^k_{j,i}.$$

Thus, $W^k_{j,i}$ represents the in-degree of vertex v_i in G^k_j and $\tilde{W}^k_{j,i}$ represents the in-degree of vertex v_i in G_j.
We will now construct these sequences. We will start with the case that $k > i$. Define $(U^k_{j,i})_{j \ge i}$ as a sequence of independent random variables on $(0, 1)$. Define

$$X^k_{i,i} = \mathbb{I}[U^k_{i,i} < 1/2i] \text{ and } \tilde{X}^k_{i,i} = \mathbb{I}[U^k_{i,i} < 1/(2i-1)].$$

Note that now

$$\mathrm{P}(X^k_{i,i}=1)=\frac{1}{2i} \text{ and } \mathrm{P}(\tilde{X}^k_{i,i}=1)=\frac{1}{2i-1}.$$

This corresponds to Equation (5.4) and (5.2) respectively.
Now, for $i<j<k$, assume that $W^k_{j-1,i}$ and $\tilde{W}^k_{j-1,i}$ are given. In other words, assume we have already constructed the first $j-1$ vertices and their connections, so we know the in-degree of v_i in G^k_j and G_j. Then, we define

$$X^k_{j,i}=\mathbb{I}\left[U^k_{j,i}<\frac{W^k_{j-1,i}}{2j}\right] \text{ and } \tilde{X}^k_{j,i}=\mathbb{I}\left[U^k_{j,i}<\frac{\tilde{W}^k_{j-1,i}}{2j-1}\right] \tag{5.6}$$

Again, this corresponds to Equation (5.4) and (5.2) respectively. Furthermore, if $j=k$, we set $X^k_{k,i}=0$, since according to our construction of G^k_n, v_k already connects to $v_{i'}\neq v_i$. If $j=k$, we set $\tilde{X}^k_{k,i}$ according to Equation (5.6). Then, for $j>k$, we define

$$X^k_{j,i}=\mathbb{I}\left[U^k_{j,i}<\frac{W^k_{j-1,i}}{2j-1}\right] \text{ and } \tilde{X}^k_{j,i}=\mathbb{I}\left[U^k_{j,i}<\frac{\tilde{W}^k_{j-1,i}}{2j-1}\right]$$

This corresponds to Equation (5.2), since now the condition $X_k=1$ has no effect anymore if adding new vertices.
We have now constructed two coupled sequences of random variables $X^k_{j,i}$ and $\tilde{X}^k_{j,i}$ which correspond to our construction of G^k_n explained after the proof of Lemma 5.9 and to the construction of G_n defined at the beginning of this section respectively.
One can easily prove by induction on j that $\tilde{W}^k_{j,i}\geq W^k_{j,i}$ and that $\tilde{X}^k_{j,i}\geq X^k_{j,i}$ for all j.
After that, define the indicator

$$A^k_{j,i}=\mathbb{I}[\min\{i\leq l\leq n|X^k_{l,i}\neq\tilde{X}^k_{l,i}\}=j].$$

Assume that for some $j>k$, $A^k_{j,i}=1$. Then, for $m=1,\dots,k$, $X^k_{m,i}=\tilde{X}_{m,i}$, and thus $W^k_{j-1,i}=\tilde{W}^k_{j-1,i}$ for all $j>k$, and thus $A^k_{j,i}=0$ for $j>k$. This is a contradiction, and thus $A^k_{j,i}=0$ for $j>k$. Thus, we have that

$$\begin{aligned}
\mathrm{P}(\tilde{W}^k_{n,i}\neq W^k_{n,i}) &= \mathrm{P}\left(\bigcup_{j=1}^{n}(A^k_{j,i}=1)\right) &(5.7)\\
&=\mathrm{P}((A^k_{k,i}=1)\cap(\tilde{X}^k_{k,i}=1))+\sum_{j=i}^{k-1}\mathrm{P}\left(\{A^k_{j,i}=1\}\cap\left\{\frac{W^k_{j-1,i}}{2j}<U^k_{j,i}<\frac{\tilde{W}^k_{j-1,i}}{2j-1}\right\}\right) &(5.8)\\
&\leq \mathrm{E}\tilde{X}^k_{k,i}+\sum_{j=i}^{k-1}\mathrm{P}\left(\frac{\tilde{W}^k_{j-1,i}}{2j}<U^k_{j,i}<\frac{\tilde{W}^k_{j-1,i}}{2j-1}\right), &(5.9)
\end{aligned}$$

with again $W^k_{i-1,i}:=\tilde{W}^k_{i-1,i}:=1$. After that, we will use Equation (5.7), Lemma 5.5 and the computations in the proof of Lemma 5.9.

$$\begin{aligned}
\mathrm{P}(\tilde{W}^k_{n,i}\neq W^k_{n,i}) &\leq \frac{\mathrm{E}W_{k-1,i}}{2k-1}+\sum_{j=1}^{n}\mathrm{E}W_{j-1,i}\left(\frac{1}{2j-1}-\frac{1}{2j}\right)\\
&\leq\frac{C_0}{2k-1}\left(\sqrt{\frac{k-1}{i^3}}+\sqrt{\frac{k-1}{i}}\right)+\sum_{j=1}^{n}C_j\left(\frac{1}{2j-1}-\frac{1}{2j}\right)\left(\sqrt{\frac{j-1}{i^3}}+\sqrt{\frac{j-1}{i}}\right)\\
&\leq C\left[\sqrt{\frac{k}{i}}\frac{1}{k}+\sqrt{\frac{k}{i^3}}\frac{1}{k}+\sum_{i=1}^{n}\left(\sqrt{\frac{j}{i}}\frac{1}{j^2}+\sqrt{\frac{j}{i^3}}\frac{1}{j^2}\right)\right]\leq C/i,
\end{aligned}$$

since $\sqrt{(k-1)/i^3} < \sqrt{k/i^3}$, $\sqrt{(k-1)/i} < \sqrt{k/i}$ and $1/(2k-1) \leq 1/(2k)$ and choose $C \geq \max\{1, C_0, C_1, ..., C_n\}$.
For $k = i$, the proof is similar and it is easy to show that $\mathrm{P}(\tilde{W}^i_{n,i} \neq W^i_{n,i} \leq C/i$. These bounds do not depend on k, so we also have that

$$\mathrm{P}(\tilde{W}^K_{n,i} - 1 \neq W^K_{n,i} - 1) \leq C/i.$$

This implies, using Lemma 4.4, we get that for p such that $\mathrm{E}W_{n,i} - 1 = (1-p)/p$, or equivalently, $p = 1/\mathrm{E}W_{n,i}$ (note that $\mathrm{E}W_{n,i} \geq 1$ and thus $0 < p \leq 1$), that for X being a random variable with the geometric distribution with rate p,

$$d_{TV}(W_{n,i}, X) \leq 2(1-p)\mathrm{P}(\tilde{W}^K_{n,i} - 1 \neq W^K_{n,i} - 1) \leq C/i,$$

since $(1-p) < 1$. ∎

We have thus obtained a bound on the total variation distance between the geometric distribution and the distribution of the in-degree of a specific node i selected from a preferential attachment graph $G^{1,0}_n$. However, remember that the main result of this section, namely Theorem 5.4, is more powerful. It namely puts a bound on the total variation distance between the Yule-Simon distribution and the distribution of the in-degree of a *random* node selected from a preferential attachment graph $G^{1,0}_n$. We will now prove this theorem.
Proof. We will first claim that for $0 \leq \epsilon \leq p$,

$$d_{TV}(Ge(p), Ge(p-\epsilon)) \leq \frac{\epsilon}{p} < \frac{\epsilon}{p-\epsilon}. \tag{5.10}$$

The last inequality is trivial. For the first inequality, consider an infinite sequence of random variables $U_0, U_1, ...$ with the continuous uniform distribution on $[0,1]$. Then, define the sequence of random variables $X_0, X_1, ...$ as $X_i = \mathbb{I}[U_i < p]$. Note that this sequence consists of i.i.d random variables with Bernoulli distribution with parameter p. Similarly, define the sequence of random variables $Y_0, Y_1, ...$ as $Y_i = \mathbb{I}[U_i < p - \epsilon]$. Note that this sequence consists of i.i.d random variables with Bernoulli distribution with parameter $p - \epsilon$. Furthermore, the sequences are constructed in such a way that there can not be a $k \in \mathbb{N}$ such that $Y_k = 1$ and $X_k = 0$. Define the random variable $X^* = \min\{k \in \mathbb{N} | X_k = 1\}$ and the random variable $Y^* = \min\{k \in \mathbb{N} | Y_k = 1\}$. Now, X^* has the geometric distribution with parameter p and Y^* has the geometric distribution with parameter $p - \epsilon$. By construction, $X^* \leq Y^*$. Thus, the probability that $X^* \neq Y^*$ is the equal to $\mathrm{P}(X^* < Y^*)$, or, equivalently, equal to the probability that for some k we have that $Y_1 = ... = Y_{k-1} = X_1 = ... = X_{k-1} = X_k = 0$ and $Y_k = 1$.

$$\begin{aligned}
\mathrm{P}(X^* < Y^*) &= \sum_{k=1}^{\infty} \mathrm{P}(Y_0 = ... = Y_{k-1} = X_0 = ... = X_{k-1} = X_k = 0, Y_k = 1) \\
&= \sum_{k=0}^{\infty} (1-p)^k \epsilon \\
&= \frac{\epsilon}{p}
\end{aligned}$$

This proves that

$$d_{TV}(Ge(p), Ge(p-\epsilon)) \leq \frac{\epsilon}{p} < \frac{\epsilon}{p-\epsilon}. \tag{5.11}$$

If we then insert $p = 1/\mathrm{E}[W_{n,i}]$ and $\epsilon = 1/\mathrm{E}[W_{n,i}] - \sqrt{i/n}$ in Equation (5.10) for any $1 \leq i \leq n$ and with $W_{n,i}$ the total degree of vertex i in the preferential attachment graph on n vertices as before, we obtain

$$d_{TV}\left(Ge\left(\frac{1}{\mathrm{E}W_{n,i}}\right), Ge\left(\sqrt{\frac{i}{n}}\right)\right) < \frac{|\frac{1}{\mathrm{E}W_{n,i}} - \sqrt{\frac{i}{n}}|}{\sqrt{\frac{i}{n}}} \leq \frac{\frac{C}{\sqrt{ni}}}{\sqrt{\frac{i}{n}}} = \frac{C}{i}. \tag{5.12}$$

For the second inequality we use Lemma 5.5.
We are now just one lemma away from proving our main theorem.

Lemma 5.11 *Let $W_{n,i}$ be the total degree of vertex i in the preferential attachment graph on n vertices and let I be uniform on $\{1, ..., n\}$ independent of $W_{n,i}$. Let U be uniform on $(0, 1)$ independent of $W_{n,i}$. Then, for a constant C,*

1. *Firstly,*
$$d_{TV}(W_{n,I} - 1, Ge(1/\mathrm{E}W_{n,I})) \leq \frac{C \log n}{n}$$
2. *Secondly,*
$$d_{TV}(Ge(1/\mathrm{E}W_{n,I}), Ge(\sqrt{I/n})) \leq \frac{C \log n}{n}$$
3. *Thirdly,*
$$d_{TV}(Ge\sqrt{I/n}, Ge\sqrt{U}) \leq \frac{C \log n}{n}$$

.

Proof.

1. We note that

$$\begin{aligned}
d_{TV}(W_{n,I} - 1, Ge(1/\mathrm{E}W_{n,I})) &\leq \mathrm{E}d_{TV}(W_{n,I} - 1|I, Ge(1/\mathrm{E}W_{n,I}|)|I) \\
&= \sum_{i=1}^{n} \frac{1}{n} d_{TV}(W_{n,i} - 1, Ge(1/\mathrm{E}W_{n,i})) \\
&\leq \sum_{i=1}^{n} \frac{1}{n}\frac{C}{i} \\
&\leq \frac{C \log n}{n}
\end{aligned}$$

The third line is based on Theorem 5.7. This proves the statement.

2. With use of Lemma 5.6 and Equation (5.12), we obtain that

$$\begin{aligned}
d_{TV}\left(Ge\left(\frac{1}{\mathrm{E}W_{n,I}}\right), Ge\left(\sqrt{\frac{I}{n}}\right)\right) &\leq \mathrm{E}d_{TV}\left(Ge\left(\frac{1}{\mathrm{E}W_{n,I}}\right)|I, Ge\left(\sqrt{\frac{I}{n}}\right)|I\right) \\
&= \sum_{i=1}^{n} \frac{1}{n} d_{TV}\left(Ge\left(\frac{1}{\mathrm{E}W_{n,i}}\right), Ge\left(\sqrt{\frac{i}{n}}\right)\right) \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \frac{C_0}{i} \\
&\leq \frac{C_0}{n}(\log n + 1) \\
&\leq \frac{C \log n}{n}
\end{aligned}$$

The fourth inequality is based on properties of harmonic sums. The fifth inequality holds for $n > 1$, since there is a C_1, e.g. $C_1 = 3$, such that $C_1 \log x > \log x + 1$ for $x > 1$. This proves the statement.

3. Now, we will couple U, having the uniform distribution on $(0,1)$, to I in the following way. Write $U = I/n - V$ with V uniform on $(0, 1/n)$ and independent of I. Note that indeed I/n is uniform on $\{1/n, 2/n, ..., 1\}$ and thus $I/n - V$ is uniform on $(0,1)$.
Then,

$$\begin{aligned}
d_{TV}(Ge\sqrt{U}, Ge\sqrt{I/n}) &= d_{TV}(Ge\sqrt{I/n - V}, Ge\sqrt{I/n}) \\
&\leq \mathrm{E} d_{TV}((Ge\sqrt{I/n - V}|I, V), (Ge\sqrt{I/n}|I, V)) \\
&= \sum_{i=1}^{n} \frac{1}{n} \int_{0}^{1/n} n d_{TV}(Ge\sqrt{i/n - v}, Ge\sqrt{i/n}) dv \\
&\leq \sum_{i=1}^{n} \int_{0}^{1/n} \frac{\sqrt{i/n} - \sqrt{i/n - v}}{\sqrt{i/n}} dv \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \frac{\sqrt{i/n} - \sqrt{(i-1)/n}}{\sqrt{i/n}} \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \frac{\sqrt{\frac{1}{ni}}}{\sqrt{i/n}} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{i} \leq \frac{C \log n}{n}.
\end{aligned}$$

For the first inequality we use Lemma 5.6 and for the fourth line we use equation (5.11). Furthermore, the fifth line follows from the fact that the term within the integral has its maximum at $v = 1/n$ and the interval of integration has length $1/n$. We also use the property that $\sqrt{a} - \sqrt{a-1} \leq \sqrt{\frac{1}{a}}$. This proves the statement.

∎

Now, by applying the triangle inequality to the three statements of Lemma 5.11, that for Z having the Yule-Simon distribution,

$$\begin{aligned}
d_{TV}(W_{n,I} - 1, Z) =& d_{TV}(W_{n,i} - 1, Ge\sqrt{U}) \\
\leq& d_{TV}(W_{n,i} - 1, Ge(1/\mathrm{E}W_{n,I})) + d_{TV}(Ge(1/\mathrm{E}[W_{n,I}|I]), Ge\sqrt{I/n}) \\
&+ d_{TV}(Ge\sqrt{I/n}), Ge\sqrt{U}) \\
\leq& \frac{C_1 \log n}{n} + \frac{C_2 \log n}{n} + \frac{C_3 \log n}{n} \\
=& \frac{C \log n}{n}
\end{aligned}$$

This proves Theorem 5.4.

∎

In this section we have gained a very powerful result for preferential attachment graphs $G_n^{1,0}$. In the next section, we will obtain a similar result for a more general family of preferential attachment graphs, namely $G_n^{m,0}$.

6 Negative binomial approximation applied to preferential attachment graphs

The outline of the proof in the last section is similar to the outline of the proof of the main result of this section. In this section we will obtain an upper bound on the total variation distance between a mixture distribution and the distribution of the in-degree of a randomly selected node from $G_n^{m,0}$. Considering the outline of the proof, we will again obtain three upper bounds, like we did in Lemma 5.11, and apply the triangle inequality to it. Again, the upper bound involving the preferential attachment graph will cover the largest part of the section. This section is based on [Ross, 2013]. In [Ross, 2013] the results are obtained for arbitrary δ, but we will focus on $\delta = 0$. From now on, we will refer to $G_n^{m,0}$ as G_n^m.

6.1 The main results and their meaning

The main result that will be proved in this section is the following theorem.

Theorem 6.1 *If W_n is the in-degree of a randomly chosen vertex from the preferential attachment graph $G_n^{m,0}$ and $K(m,0)$ is the mixed negative binomial distribution as defined in Definition 3.4, then for some constant C_m,*

$$d_{TV}(W_n, K(m,0)) \leq \frac{C_m \log n}{n}$$

.

In [Ross, 2013] a more general result of this theorem is proven, namely the following.

Theorem 6.2 *If W_n is the in-degree of a randomly chosen vertex form the preferential attachment graph $G_n^{m,\delta}$ and $K(m,\delta)$ is the mixed negative binomial distribution as defined in Definition 3.4, then for some constant $C_{m,\delta}$,*

$$d_{TV}(W_n, K(m,\delta)) \leq \frac{C_{m,\delta} \log n}{n}$$

.

This is a very powerful result, because of the power law behaviour of $K(m,\delta)$, which is stated in the following lemma.

Lemma 6.3 *If $m \in \mathbb{N}$, $\delta > -m$ and $Z \sim K(m,\delta)$, then, for $l = 0, 1, \ldots$,*

$$\mathrm{P}(Z = l) = \left(2 + \frac{\delta}{m}\right) \frac{\Gamma(l+m+\delta)\Gamma(m+2+\delta+\frac{\delta}{m})}{\Gamma(m+\delta)\Gamma(l+m+3+\delta+\frac{\delta}{m})},$$

and for $C_{m,\delta} = (2+\delta/m)\Gamma(m+2+\delta+\frac{\delta}{m})/\Gamma(m+\delta)$,

$$\mathrm{P}(Z = k) \to \frac{C_{m,\delta}}{k^{3+\delta/m}}$$

as $k \to \infty$.

We will prove this lemma for the case that we are going to treat, namely $\delta = 0$.
However, first we need another result.

Lemma 6.4 *For fixed $a, b > 0$ as $z \to \infty$,*

$$\frac{(z+a)!}{(z+b)!} = z^{a-b} + O\left(z^{a-b-1}\right)$$

Proof. We will prove this lemma with use of Stirling's approximation, which states that for $n \to \infty$,

$$n \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Thus, we have that for $z \to \infty$,

$$\frac{(z+a)!}{(z+b)!} \approx \sqrt{\frac{z+a}{z+b}} e^{b-a} \frac{(z+a)^{z+a}}{(z+b)^{z+b}}$$

We have that $\sqrt{\frac{z+a}{z+b}} \approx 1$ and

$$\begin{aligned}\frac{(z+a)^{z+a}}{(z+b)^{z+b}} &= \frac{(z+a)^{z+a}}{(z+b)^{z+a}}(z+b)^{a-b} \\ &= \left(1 + \frac{a-b}{z+b}\right)^{z+a} (z+b)^{a-b} \\ &\approx e^{(a-b)(z+a)/(z+b)}(z+b)^{a-b} \\ &\approx e^{a-b}(z+b)^{a-b}\end{aligned}$$

This yields that

$$\frac{(z+a)!}{(z+b)!} \approx (z+b)^{a-b},$$

which gives the desired result. ∎

We will now prove Lemma 6.3.

Proof.

For $m \in \mathbb{N}$, $Z \sim K(m, 0)$ and U uniform on $(0, 1)$, we have to show that

$$\mathrm{P}(Z = l) = 2\frac{\Gamma(l+m)\Gamma(m+2)}{\Gamma(m)\Gamma(l+m+3)} = 2\frac{(l+m-1)!(m+1)!}{(m-1)!(l+m+2)!}.$$

We have that

$$\begin{aligned}\mathrm{P}(Z = l) =& \mathrm{P}(NB(m, \sqrt{U}) = l) \\ =& \int_0^1 \mathrm{P}(NB(m, \sqrt{u}) = l) du \\ =& \int_0^1 \binom{l+m-1}{l} (1-\sqrt{u})^l \sqrt{u}^m du \\ =& \binom{l+m-1}{l} \int_0^1 (1-\sqrt{u})^l \sqrt{u}^m du.\end{aligned}$$

We will solve the integral with use of the following property, which we obtain by integration by parts. For $k, n \in \mathbb{N}$,

$$\begin{aligned} I_{k,n} &:= \int_0^1 (1-\sqrt{u})^k \sqrt{u}^n du \\ &= \left[(1-\sqrt{u})^k \frac{2}{n+2}\sqrt{u}^{n+2}\right]_{u=0}^{u=1} - \int_0^1 -\frac{k}{2}u^{-1/2}(1-\sqrt{u})^{k-1}\frac{2}{n+2}\sqrt{u}^{n+2}du \\ &= \frac{k}{n+2}\int_0^1 (1-\sqrt{u})^{k-1}\sqrt{u}^{n+1}du = \frac{k}{n+2}I_{k-1,n+1} \end{aligned}$$

Thus, we get that

$$I_{k,n} = \prod_{i=1}^{k} \frac{k-i+1}{n+i+1} I_{0,n+k},$$

and we calculate that

$$I_{0,n+k} = \int_0^1 \sqrt{u}^{n+k} du = \frac{2}{n+k+2}$$

Thus, combining the results,

$$I_{k,n} = \frac{2}{n+k+2}\prod_{i=1}^{k}\frac{k-i+1}{n+i+1}.$$

We get that

$$\begin{aligned} \mathrm{P}(Z=l) =& \binom{l+m-1}{l} I_{l,m} \\ =& \binom{l+m-1}{l}\frac{2}{l+m+2}\prod_{i=1}^{l}\frac{l-i+1}{m+i+1} \\ =& \frac{(l+m-1)!}{l!(m-1)!}\frac{2}{l+m+2}\frac{l!(m+1)!}{(l+m+1)!} \\ =& 2\frac{(l+m-1)!(m+1)!}{(m-1)!(l+m+2)!}, \end{aligned}$$

which is our desired result. Furthermore, we get from Lemma 6.4 with $z = k$, $a = m-1$, $b = m+2$ that for $k \to \infty$

$$\mathrm{P}(Z=k) = 2\frac{(m+1)!}{(m-1)!}\frac{(k+m-1)!}{(k+m-1)!} \asymp \frac{C}{l^3}.$$

This shows the second statement.

∎

Theorem 6.2 and Lemma 6.3 illustrate that the random variable of the in-degree of a randomly selected node from a preferential attachment graph behaves according to a power-law. This is a very special property, since the process of preferential attachment is not constructed to resemble the power law behaviour that is often seen in real-world networks. Thus, we could say that preferential attachment graphs have a high explanatory value considering the power law behaviour in real-world networks.

6.2 Proof of the main results

We need many intermediate results before we are ready to prove Theorem 6.1.

We will start with an analogous lemma to Lemma 5.5, which is proven in [Ross, 2013].

Lemma 6.5 *Let $W_{n,i}^m$ be the in-degree of vertex i in the preferential attachment graph G_n^m. Then,*

$$\left|\frac{\mathrm{E}W_{n,i}^m}{m}+1-\sqrt{\frac{n}{i}}\right| \leq C_m\sqrt{\frac{n}{i^3}} \text{ and } \left|\frac{m}{\mathrm{E}W_{n,i}^m+m}-\sqrt{\frac{i}{n}}\right| \leq \frac{C_m}{\sqrt{ni}}$$

with C_m a constant dependent on m.

We will now prove the following lemma, which is very similar to Equation 5.11 in the proof of Theorem 5.4.

Lemma 6.6 *If $r \in \mathbb{N}$ and $0 \leq \epsilon \leq p$, then*

$$d_{TV}(NB(r,p), NB(r,p-\epsilon)) \leq \frac{r\epsilon}{p-\epsilon}$$

Proof.

Consider an infinite sequence of random variables $U_0, U_1, ...$ with the continuous uniform distribution on $[0,1]$. Then, define the sequence of random variables $X_0, X_1, ...$ as $X_i = \mathbb{I}[U_i < p]$. Note that this sequence consists of i.i.d random variables with Bernoulli distribution with parameter p. Similarly, define the sequence of random variables $Y_0, Y_1, ...$ as $Y_i = \mathbb{I}[U_i < p - \epsilon]$. Note that this sequence consists of i.i.d random variables with Bernoulli distribution with parameter $p - \epsilon$. Furthermore, the sequences are constructed in such a way that there can not be a $k \in \mathbb{N}$ such that $Y_k = 1$ and $X_k = 0$. Define the random variable $X^* = \min\{k \in \mathbb{N} | \exists i_1, ..., i_r \leq k \text{ s.t. } X_{i_1} = \cdots = X_{i_r} = 1\}$ and the random variable $Y^* = \min\{k \in \mathbb{N} | \exists j_1, ..., j_r \leq k \text{ s.t. } Y_{j_1} = \cdots = Y_{j_r} = 1\}$. Now, X^* has the negative binomial distribution with parameters r and p and Y^* has the negative binomial distribution with parameters r and $p - \epsilon$.
By construction, $X^* \leq Y^*$. Thus, the probability that $X^* \neq Y^*$ is the equal to $\mathrm{P}(X^* < Y^*)$, or, equivalently, equal to the probability that for some combination $i_1 < \cdots < i_r$ such that $Y_{i_1} = \cdots = Y_{i_r} = 1$ and $Y_i = 0$ for $i \leq i_r, i \neq i_1, ..., i_r$, there exists a $k \in \{i_1, ..., i_r\}$ such that $X_k = 0$. Thus,

$$\begin{aligned}
\mathrm{P}(X^* < Y^*) = &\sum_{i_1<\cdots<i_r}\sum_{k=1}^{r}\mathrm{P}(X_{i_1} = \cdots = X_{i_{k-1}} = 1, X_{i_k} = 0 | Y_{i_1} = \cdots Y_{i_r} = 1) \\
&\cdot \mathrm{P}(Y_{i_1} = \cdots = Y_{i_r} = 1; Y_i = 0 \text{ for } i \leq i_r, i \neq i_1, ..., i_r) \\
=&\sum_{k=1}^{r}\left(\frac{p-\epsilon}{p}\right)^{k-1}\frac{\epsilon}{p}\sum_{i_1<\cdots<i_r}\mathrm{P}(Y_{i_1} = \cdots = Y_{i_r} = 1; Y_i = 0 \text{ for } i \leq i_r, i \neq i_1, ..., i_r) \\
=&\frac{\epsilon}{p}\frac{1-\left(\frac{p-\epsilon}{p}\right)^r}{1-\frac{p-\epsilon}{p}} = 1-\left(1-\frac{\epsilon}{p}\right)^r
\end{aligned}$$

Note that after the second equality, the first summation is recognized as a geometric sum and the second summation equals the probability that somewhere in the infinite sequence of X_ns, the r^{th} success occurs. This probability equals 1. Now we are going to show that

$$P(X^* < Y^*) = 1 - \left(1 - \frac{\epsilon}{p}\right)^r \leq \frac{r\epsilon}{p}.$$

Note that $1 - \left(1 - \frac{\epsilon}{p}\right)^r$ equals the probability of at least one success among r i.i.d. Bernoulli random variables with probability of success $\frac{\epsilon}{p}$. This probability is smaller than r times the probability of success. This explains that

$$P(X^* < Y^*) = 1 - \left(1 - \frac{\epsilon}{p}\right)^r \leq \frac{r\epsilon}{p}.$$

This proves our statement and thus,

$$d_{TV}(NB(r,p), NB(r, p-\epsilon)) \leq \frac{r\epsilon}{p} < \frac{r\epsilon}{p-\epsilon}.$$

∎

We will now introduce a theorem that constitutes an important result of this section and is analogous to Theorem 5.7.

Theorem 6.7 *Let $W^m_{n,i}$ be the in-degree of vertex i in G^m_n. Then, we have that*

$$d_{TV}\left(W^m_{n,i}, NB\left(m, \frac{m}{\mathrm{E}W^m_{n,i} + m}\right)\right) \leq \frac{C_m}{i},$$

where C_m is a constant depending on m.

Note that

$$\mathrm{E}NB\left(m, \frac{m}{\mathrm{E}W^m_{n,i} + m}\right) = \frac{m\left(1 - \frac{m}{\mathrm{E}W^m_{n,i}+m}\right)}{\frac{m}{\mathrm{E}W^m_{n,i}+m}} = \mathrm{E}W^m_{n,i}.$$

Proof. We will use the framework suggested by Theorem 3.9 to prove this theorem, and thus we will construct a random variable having the m-equilibrium distribution with respect to $W^m_{n,i}$. Thus, we first need a random variable having the size-biased distribution with respect to $W^m_{n,i}$. Since the construction of G^m_n makes use of the construction of G^1_{mn}, we will again write, for $k \geq j$ and $W^1_{k,j}$ being the in-degree of vertex j in G^1_k,

$$W^m_{n,i} = \sum_{j=1}^{m} W^1_{mn,m(i-1)+j}.$$

Furthermore, let $X^1_{j,s}$ be the indicator that vertex j attaches to vertex g in G^1_j, and thus also in G^1_k for $j \leq k \leq mn$. Then, we have that

$$W^1_{mn,m(i-1)+j} = \sum_{k=m(i-1)+j}^{mn} X^1_{k,m(i-1)+j}.$$

We will again use Proposition 5.8 to construct a random variable with the size-biased distribution with respect to $W^m_{n,i}$. In the proposition, the random variable with the size-biased distribution of $X = \sum_{j=1}^{n} X_j$ is constructed. We want to construct a random variable with the size-biased distribution of

$$W_{n,i}^m = \sum_{j=1}^{m} \sum_{k=m(i-1)+j}^{mn} X_{k,m(i-1)+j}^1 = \sum_{l=m(i-1)+1}^{mi} \sum_{k=l}^{mn} X_{k,l}^1.$$

This suggests that we define $p_{k,l} = \mathrm{P}(X_{k,l}^1 = 1)$. Note that for $k < l$, $p_{k,l} = \mathrm{P}(X_{k,l}^1 = 1) = 0$ Furthermore, we define the integer-valued random variables K and L such that for $l = m(i-1)+1, ..., mi$ and $k = l, ..., mn$, $\mathrm{P}(K = k, L = l)$ is proportional to $p_{k,l} = \mathrm{P}(X_{k,l}^1 = 1)$ and zero for all other values. Note that this statement is satisfied and the probabilities are properly defined for

$$p_{k,l} = \frac{\mathrm{E}X_{k,l}^1}{\mathrm{E}W_{n,i}^m}.$$

Thus, to size-bias $W_{n,i}^m$, we first sample L and K according to the probabilities defined above. Note that L determines which one of the m vertices in G_{mn}^1 that are used to construct vertex i in G_n^m is considered and K determines which one of the later added vertices in G_{mn}^1 will connect to this vertex in G_{mn}^i. Assume $L = l, K = k$. After that, we introduce the condition that $X_{k,l}^1 = 1$. In other words, we add an outgoing edge from v_k to v_l. Because of the values that l can obtain, this is an incoming edge of vertex i in G_n^m. Then, we sample $X_{j,i}^{(k,l)}$ for $m(i-1)+1 \le i \le mi, l \le k \le mn$ with $(j,i) \ne (k,l)$. I.e. we sample the indicator variable $X_{j,i}^1$ with the condition that $X_{k,l}^1 = 1$. Then, we count the number of vertices connecting to $v_{m(i-1)+1}, ..., v_{mi}$. This equals sum of the sampled values of $X_{j,s}$ with $(j,s) \ne (k,l)$ plus 1 since $X_{k,l} = 1$.
For $j < l$ and $j > k$, this does not change the rule for generating the preferential attachment graph given G_{j-1}^1. For $l \le j < k$ we have a result similar to Lemma 5.9.

Lemma 6.8 *Retaining the notation and definitions above, for $l, s \le j < k$, we have*

$$\mathrm{P}(X_{j,s}^1 = 1 | X_{k,l}^1 = 1, G_{j-1}^1) = \frac{\mathbb{I}[s = l] + W_{j-1,s}^1 + 1}{2j}$$

Note that the meaning of this lemma is equivalent to the meaning of Lemma 5.9. I.e., the process of generating the graphs G_j^1 for $l \le j < k$ conditional on $X_{k,l} = 1$ and G_{l-1} is equivalent to our normal rule for preferential attachment, but we have to include the edge from vertex k to vertex l in the degree count. Indeed, the numerator equals the total degree of vertex s in G_{j-1}^1 plus 1 if $s = l$, i.e. plus 1 if vertex k connects to vertex s. The denominator equals the total degree of the vertices from regular preferential attachment plus 1 because of the incoming edge from vertex k to vertex l.
The proof of this lemma is very similar to the proof of Lemma 5.9, so it won't be discussed here. For details, we refer to [Ross, 2013].
This lemma suggests the following construction of $(W_{n,i}^m | X_{k,l} = 1)$. We denote variables relating to this construction by an extra superscript k, l. First, we generate $G_{l-1}^{1,k,l}$ by the usual preferential attachment model. If $l \ne k$, v_l and v_k are added to the graph, along with a vertex $v_{i'}$ with an edge to it emanting from v_k. Given $G_{l-1}^{1,k,l}$ and the additional vertices and edges, we generate $G_l^{1,k,l}$ by connecting vertex l to a vertex randomly chosen from $v_1, ..., v_l, v_{i'}$ with probability proportional to their degree. Note that v_l has degree 1 from its out-degree and vertex $v_{i'}$ has degree 1 from its in-degree. For $l < j < k$, we generate the graphs $G_j^{1,k,l}$ recursively from $G_{j-1}^{1,k,l}$ by connecting v_j to a vertex randomly chosen from the vertices $v_1, ..., v_j, v_{i'}$ with

probability proportional to their degree weight again. Note that in this process, $v_1, ..., v_{k-1}$ can not connect to v_k, but can connect to $v_{i'}$. Furthermore, define $G_k^{1,k,l} = G_{k-1}^{1,k,l}$, since v_k has already 'used' its out-degree in connecting to $v_{i'}$. If $l = k$ we attach v_k to $v_{i'}$ and denote the resulting graph by $G_k^{1,k,l}$. For all values $j = k+1, ..., mn$ we generate $G_j^{1,k,l}$ from $G_{j-1}^{1,k,l}$ according to usual preferential attachment among $v_1, ..., v_j, v_{i'}$. Denote the total degree of vertex j in this construction by $W_{mn,j}^{1,k,l}$ and let also

$$W_{n,i}^{m,k,l} = \sum_{j=1}^{m} W_{mn,m(i-1)+j}^{1,k,l}$$

After that, let $B_{k,l}$ be the event that in this construction, all edges emanting from the vertices $m(i-1)+1, ..., mi$ attach to one of the vertices $1, ..., m(i-1)$. I.e., the event that vertex i in G_n^m does not connect to itself or to $v_{i'}$.

$$\begin{aligned} B_{k,l} = \{ & X_{j,s}^{k,l} = 0 \text{ for all } s \in \{m(i-1)+1, ..., mi, i'\}, \\ & j \in \{m(i-1)+1, ..., mi, i'\} \backslash \{k\}\} \end{aligned}$$

After that, let W' have the m-equilibrium distribution with respect to $W_{n,i}^m$.
Note that we do not have any information on this distribution, but in the continuation of this proof, we note that we do not need this.
Furthermore, define

$$W_{n,i}^{*m} = W_{n,i}^{m,K,L} \mathbb{I}[B_{K,L}] + W' \mathbb{I}[B_{K,L}^c]. \tag{6.1}$$

By $B_{K,L}^c$ we mean the complement of $B_{K,L}$.
The next lemma combines the obtained results to construct a random variable having the m-equilibrium distribution with respect to $W_{n,i}^m$ and is similar to Lemma 5.10.

Lemma 6.9 *Let $l \in \{m(i-1)+1, ..., mi\}$, $k \in \{l, ..., mn\}$ and retain the notation and definitions above. Then,*

1. $\mathcal{L}(W_{n,i}^{m,k,l} + W_{mn,i'}^{m,k,l}) = \mathcal{L}(W_{n,i}^m | X_{k,l}^1 = 1)$
2. *If (K, L) is a random vector such that*

   $$\mathrm{P}(K = k', L = l') = \frac{\mathrm{E}X_{k',l'}^1}{\mathrm{E}W_{n,i}^m}, k' \geq l \in \{m(i-1)+1, ..., mi\},$$

   then $W_{n,i}^{m,K,L} + W_{mn,i'}^{1,K,L}$ has the size-bias distribution with respect to $W_{n,i}^m$.
3. *Conditional on the event*

   $$\{W_{n,i}^{m,K,L} + W_{mn,i'}^{1,K,L} = t\},$$

   $$\mathcal{L}(W_{n,i}^{m,K,L} \mathbb{I}[B_{K,L}]) = \mathcal{L}(U_{m,t} \mathbb{I}[B_{K,L}]),$$

   where $U_{r,t}$ has the Pólya urn distribution as described in Subsection 3.2 and is independent of all else.
4. *$W_{n,i}^{*m}$ has the m-equilibrium distribution of $W_{n,i}$*

Proof.

1. This statement follows from the construction of $W_{n,i}^{m,k,l}$ and $W_{mn,i'}^{1,k,l}$.
2. Using Proposition 5.8 and Lemma 6.8, we obtain this result.

3. If $\mathbb{I}[B_{K,L}] = 1$, conditional on $X^1_{K,L} = 1$, all edges emanating from the vertices $v_{m(i-1)+1}, ..., v_{mi}$ attach to one of the vertices $v_1, ..., mi-1$. Thus, in G_{mi}, $v_{m(i-1)+1}, ..., v_{mi}$ have in-degree 1, or, collapsing the vertices, v^m_i has out-degree m. Furthermore, $v^m_{i'}$ has in-degree 1. Thus, since $\{W^{m,K,L}_{n,i} + W^{1,K,L}_{mn,i'} = t\}$ we know that $t-1$ of the vertices $v_{mi+1}, ..., v_{mn}$ will connect to a vertex in $\{v_{m(i-1)+1}, ..., v_{mi}\}$ or to $v_{i'}$. Note that thus, $W^{m,K,L}_{n,i}$, or how many out of the $t-1$ connections connect to $\{v_{m(i-1)+1}, ..., v_{mi}\}$ is distributed as a Pólya urn starting with m white balls (the total degree of $\{v_{m(i-1)+1}, ..., v_{mi}\}$ in $G^{1,K,L}_{mi}$) and 1 black ball (the total degree of $v_{i'}$ in $G^{1,K,L}_{mi}$). This proves the statement.

4. This statement follows from the definition of the m-equilibrium distribution (Definition 3.6). Namely, note that X^{*m} has the m-equilibrium distribution with respect to $W^m_{n,i}$ if, for X^s having the size-bias distribution with respect to $W^m_{n,i}$, for $k = 0, 1, ...$,

$$
\begin{aligned}
\mathrm{P}(X^{*m} = k) =& \mathrm{P}(U_{r,X^s} = k) && (6.2)\\
=& \mathrm{P}(U_{r,W^{m,K,L}_{n,i} + W^{1,K,L}_{mn,i'}} = k) && (6.3)\\
=& \sum_{t=1}^{m(n-i+1)} \mathrm{P}(W^{m,K,L}_{n,i} + W^{1,K,L}_{mn,i'} = t)\mathrm{P}(U_{r,t} = k) && (6.4)\\
=& \sum_{t=1}^{m(n-i+1)} [\mathrm{P}(W^{m,K,L}_{n,i} + W^{1,K,L}_{mn,i'} = t)\mathrm{P}(U_{r,t} = k)\mathrm{P}(B_{K,L} = 1) && (6.5)\\
& + \mathrm{P}(W^{m,K,L}_{n,i} + W^{1,K,L}_{mn,i'} = t)\mathrm{P}(U_{r,t} = k)\mathrm{P}(B^c_{K,L} = 1)] && (6.6)\\
=& \sum_{t=1}^{m(n-i+1)} \mathrm{P}(W^{m,K,L}_{n,i} = t)\mathrm{P}(U_{r,t} = k)\mathrm{P}(B_{K,L} = 1) + \mathrm{P}(W' = k)\mathrm{P}(B^c_{K,L} = 1) && (6.7)\\
=& \sum_{t=1}^{m(n-i+1)} \mathrm{P}(W^{m,K,L}_{n,i} = t)\mathrm{P}(W^{m,K,L}_{n,i} = k)\mathrm{P}(B_{K,L} = 1) + \mathrm{P}(W' = k)\mathrm{P}(B^c_{K,L} = 1) && (6.8)\\
=& \left[\sum_{t=1}^{m(n-i+1)} \mathrm{P}(W^{m,K,L}_{n,i} = t)\right] \mathrm{P}(W^{m,K,L}_{n,i} = k)\mathrm{P}(B_{K,L} = 1) + \mathrm{P}(W' = k)\mathrm{P}(B^c_{K,L} = 1) && (6.9)\\
=& \mathrm{P}(W^{m,K,L}_{n,i} = k)\mathrm{P}(B^c_{K,L} = 1) + \mathrm{P}(W' = k)\mathrm{P}(B^c_{K,L} = 1) && (6.10)\\
=& \mathrm{P}(W^{m,K,L}_{n,i}\mathbb{I}[B_{K,L}] + W'\mathbb{I}[B^c_{K,L}] = k). && (6.11)
\end{aligned}
$$

Equation (6.3) follows from Definition 3.6, Equation (6.4) follows from statement 2, Equation (6.6) follows from $\mathbb{I}[B_{K,L}] + \mathbb{I}[B^c_{K,L}] = 1$, Equation (6.7) follows from the definition of W' and statement 2, Equation (6.8) follows from statement 3, Equation (6.10) follows from

$$\sum_{t=1}^{m(n-i+1)} \mathrm{P}(W^{m,K,L}_{n,i} = t) = 1$$

and Equation (6.11) follows from $\mathbb{I}[B_{K,L}] + \mathbb{I}[B^c_{K,L}] = 1$. This proves the statement.

∎

We are now ready to prove Theorem 6.7, by bounding

$$d_{TV}\left(W^m_{n,i}, NB\left(m, \frac{m}{\mathrm{E}W^m_{n,i}+m}\right)\right).$$

Note that Theorem 3.9 states that for $W^m_{n,i}$ and X^{*m} a random variable with the m-equilibrium distribution with respect to X,

$$d_{TV}\left(W^m_{n,i}, NB\left(m, \frac{m}{\mathrm{E}W^m_{n,i}+m}\right)\right) \leq 2(e\max\{1,m\}+1)\mathrm{P}(X^{*m} \neq W^m_{n,i}) = 2e(m+1)\mathrm{P}(X^{*m} \neq W^m_{n,i}).$$

We know that

$$\begin{aligned}\mathrm{P}(X^{*m} \neq W^m_{n,i}) =& \mathrm{P}(W^{m,K,L}_{n,i}\mathbb{I}[B_{K,L}] + W'\mathbb{I}[B^c_{K,L}] \neq W^m_{n,i}) \\ =& \mathrm{P}(W^{m,K,L}_{n,i} \neq W^m_{n,i}, B_{K,L}) + \mathrm{P}(W' \neq W^m_{n,i}, B^c_{K,L}) \\ \leq& \mathrm{P}(W^{m,K,L}_{n,i} \neq W_{n,i}) + \mathrm{P}(B^c_{K,L}).\end{aligned}$$

Note that here we lose the need for information about the unknown random variable W'. We will start by bounding the first term. For $l \in \{m(i-1)+1, ..., mi\}$ and $k > mi$, and $v_j \to \{v_1, ..., v_s\}$ describing the event that v_j connects to a vertex in the set $\{v_1, ..., v_s\}$.

$$\begin{aligned}\mathrm{P}(B_{k,l}) =& \prod_{j=m(i-1)+1}^{mi} \mathrm{P}(v_j \to \{v_1, ..., v_{m(i-1)}\}) \\ =& \prod_{j=m(i-1)+1}^{l-1} \frac{m(i-1)+j-1}{2j-1} \prod_{j=l}^{mi} \frac{m(i-1)+j-1}{2j} \\ \geq& \prod_{j=m(i-1)+1}^{mi} \frac{m(i-1)+j-1}{2j} \\ =& 2^{-m}\frac{(m(i-1))!(2m(i-1)+m-1)!}{(mi)!(2m(i-1)-1)!} \\ =& 1 + O(1/i)\end{aligned}$$

In the second equality the numerators in both products describe the total weight of $\{v_1, ..., v_{m(i-1)}\}$ in $G^{1,k,l}_j$ if all vertices in $\{v_{m(i-1)+1}, ..., v_j\}$ connect to a vertex in $\{v_1, ..., v_{m(i-1)}\}$. Furthermore, the denominators in both products describe the total weight in $G^{1,k,l}_j$. Note that this total weight is 1 higher for $j \geq l$, since the connection from v_k to v_l is then considered. The last equality follows from the property that

$$\frac{(z-1+a)!}{(z-1+b)!} \to z^{a-b} + O(z^{a-b-1})$$

as $z \to \infty$. This result follows from Stirling's approximation.
Furthermore, if $k \in \{m(i-1)+1, ..., mi\}$,

$$\mathrm{P}(B_{k,l}) = \prod_{\substack{j=m(i-1)+1 \\ j\neq k}}^{mi} \mathrm{P}(v_j \to \{v_1, ..., v_{m(i-1)}\}),$$

which is greater than

$$2^{-m}\frac{(m(i-1))!(2m(i-1)+m-1)!}{(mi)!(2m(i-1)-1)!},$$

since the omitted factor is a probability. This means that

$$\mathrm{P}(B^c_{K,L}) = O(1/i).$$

To bound $\mathrm{P}(W^{m,K,L}_{n,i} \neq W_{n,i})$, we must couple the random variable $W^m_{n,i}$ to $W^{m,K,L}_{n,i}$. This process is comparable to what we did in the proof of Theorem 5.7. For each (k,l) such that $l = m(i-1)+1, ..., mi$ and $k = l, ..., mn$, i.e. for each (k,l) in the support of (K,L) we construct the sequences of random variables

$$\left\{\left(X^{1,k,l}_{s,j}, \tilde{X}^{1,k,l}_{s,j}\right) | mn \geq s \geq j \in \{m(i-1)+1, ..., mi\}\right\}$$

with $X^{1,k,l}_{s,j}$ having the distribution of the indicator of the event that v_s connects to v_j in $G^{1,k,l}_{mn}$ and $\tilde{X}^{1,k,l}_{s,j}$ having the distribution of the indicator of the event that v_s connects to v_j in G^1_{mn}. Then, we can denote

$$W^{1,k,l}_{mn,j} = \sum_{s=j}^{mn} X^{1,k,l}_{s,j} \text{ and } \tilde{W}^{1,k,l}_{mn,j} = \sum_{s=j}^{mn} \tilde{X}^{1,k,l}_{s,j},$$

which have the distribution of the weight of vertex j in $G^{1,k,l}_{mn}$ and G^1_{mn} respectively. Then, we set

$$W^{m,k,l}_{n,i} = \sum_{j=m(i-1)+1}^{mi} W^{1,k,l}_{mn,j} \text{ and } W^m_{n,i} = \sum_{j=m(i-1)+1}^{mi} \tilde{W}^{1,k,l}_{mn,j}.$$

Then, we get that

$$\begin{aligned}\mathrm{P}(W^{m,k,l}_{n,i} \neq W_{n,i}) \leq & \mathrm{P}\left(\bigcup_{j=m(i-1)+1}^{mi} \left\{W^{1,k,l}_{mn,j} \neq \tilde{W}^{1,k,l}_{mn,j}\right\}\right) \\ \leq & \sum_{j=m(i-1)+1}^{mi} \mathrm{P}(W^{1,k,l}_{mn,j} \neq \tilde{W}^{1,k,l}_{mn,j}),\end{aligned}$$

and we show that each term of the sum is $O(i/1)$, which establish the theorem.
First, assume $j < l < k$. Let $U^{k,l}_{s,j}$ be uniform random variables on $(0,1)$. First, define

$$X^{1,k,l}_{j,j} = \mathbb{I}\left[U^{k,l}_{s,j} < \frac{1}{2j-1}\right]$$

and for $j < s < l$, given $W^{1,k,l}_{s-1,j}$,

$$X^{1,k,l}_{s,j} = \mathbb{I}\left[U^{k,l}_{s,j} < \frac{W^{1,k,l}_{s-1,j}+1}{2s-1}\right]$$

Furthermore, for $j \leq s < l$ we set $\tilde{X}^{1,k,l}_{s,j} = X^{1,k,l}_{s,j}$. Note that indeed the process of creating $G^{1,k,l}_s$ is the same to the process of creating G^1_s until v_l arrives. Now, for $l \leq s < k$, given $W^{1,k,l}_{s-1,j}$ and $\tilde{W}^{1,k,l}_{s-1,j}$, define

$$X^{1,k,l}_{s,j} = \mathbb{I}\left[U^{k,l}_{s,j} < \frac{W^{1,k,l}_{s-1,j}+1}{2s}\right]$$

and

$$\tilde{X}^{1,k,l}_{s,j} = \mathbb{I}\left[U^{k,l}_{s,j} < \frac{\tilde{W}^{1,k,l}_{s-1,j}+1}{2s-1}\right]. \tag{6.12}$$

Note that the numerator of the first fraction has increased by 1, because of the extra connection from v_k to $v_{i'}$. Furthermore, set $X_{k,j}^{1,k,l} = 0$, since v_k is already connected to the special vertex $v_{i'}$ and set $\tilde{X}_{k,j}$ as in Equation (6.12) with $s = k$. For $s > k$, define

$$X_{s,j}^{1,k,l} = \mathbb{I}\left[U_{s,j}^{k,l} < \frac{W_{s-1,j}^{1,k,l} + 1}{2s - 1}\right]$$

and

$$\tilde{X}_{s,j}^{1,k,l} = \mathbb{I}\left[U_{s,j}^{k,l} < \frac{\tilde{W}_{s-1,j}^{1,k,l} + 1}{2s - 1}\right].$$

It is clear that the constructed random variables correspond to the construction of the graphs $G_{mn}^{1,k,l}$ and G_{mn}^1 and that $W_{mn,j}^{1,k,l}$ and $\tilde{W}_{mn,j}^{1,k,l}$ are indeed the required degree counts. We can easily prove by induction on s that $\tilde{W}_{s,j}^{1,k,l} \geq W_{s,j}^{1,k,l}$ and $\tilde{X}_{s,j}^{1,k,l} \geq X_{s,j}^{1,k,l}$ for all s. Now define

$$A_{s,j}^{k,l} = \mathbb{I}\left[\min\left\{j \leq t \leq mn | X_{t,j}^{1,k,l} \neq \tilde{X}_{t,j}^{1,k,l}\right\} = s\right].$$

We use that $\tilde{W}_{s-1,j}^{1,k,l} = W_{s-1,j}^{1,k,l}$ if $A_{s,j}^{k,l} = 1$. Furthermore, if $\tilde{X}_{s,j}^{k,l} = X_{s,j}^{k,l}$ for all $s \leq k$, $\tilde{W}_{k,j}^{k,l} = W_{k,j}^{k,l}$ and thus $\tilde{X}_{s,j}^{k,l} = X_{s,j}^{k,l}$ for all $s > k$. This means that $A_{s,j}^{k,l} = 0$ for all $s > k$. Using these two properties, we find that

$$\begin{aligned}
\mathrm{P}\left(W_{mn,j}^{1,k,l} \neq \tilde{W}_{mn,j}^{1,k,l}\right) &= \mathrm{P}\left(\bigcup_{s=j}^{mn} A_{s,j}^{k,l}\right) \\
&\leq \sum_{s=l}^{k} \mathrm{P}\left(A_{s,j}^{k,l} \bigcap \left\{\frac{W_{s-1,j}^{1,k,l} + 1}{2s} < U_{s,j}^{k,l} < \frac{\tilde{W}_{s-1,j}^{1,k,l} + 1}{2s - 1}\right\}\right) \\
&\leq \mathrm{E}\tilde{X}_{k,j}^{1,k,m} + \sum_{s=l}^{k-1} \mathrm{P}\left(\frac{\tilde{W}_{s-1,j}^{1,k,l} + 1}{2s} < U_{s,j}^{k,l} < \frac{\tilde{W}_{s-1,j}^{1,k,l} + 1}{2s - 1}\right) \\
&\leq \frac{\mathrm{E}W_{k-1,j}^{1} + 1}{2k - 1} + \sum_{s=l}^{k-1} (\mathrm{E}W_{s-1,j}^{1} + 1)\left(\frac{1}{2s - 1} - \frac{1}{2s}\right)
\end{aligned}$$

Now, we use the bounds of Lemma 6.5 to obtain

$$\begin{aligned}
\mathrm{P}\left(W_{mn,j}^{1,k,l} \neq \tilde{W}_{mn,j}^{1,k,l}\right) \leq & \frac{C_k}{2k - 1}\sqrt{\frac{k-1}{j^3}} + \frac{1}{2k - 1}\sqrt{\frac{k-1}{j}} \\
& + \sum_{s=l}^{k-1} C_s \left(\frac{1}{2s - 1} - \frac{1}{2s}\right)\left(\sqrt{\frac{s-1}{j^3}} + \sqrt{\frac{s-1}{j}}\right) \leq C/j
\end{aligned}$$

For $l < j < k$ the coupling is similar, except that we don't have to cope with the first part of the construction, where $\tilde{X}_{s,j}^{1,k,l} = \tilde{X}_{s,j}^{1,k,l}$ and can start with the part where $j \leq s < k$. If $j > k$, the variables can be perfectly coupled. If $j = k$ or $j < l = k$, the coupling can only differ if the edge emanating from v_k connects to v_j in G_k^1, which occurs with order $O(1/i)$. Thus, eventually, we can say that

$$\mathrm{P}(W_{n,i}^{m,k,l} \neq W_{n,i}) \leq \sum_{j=m(i-1)+1}^{mi} \mathrm{P}(W_{mn,j}^{1,k,l} \neq \tilde{W}_{mn,j}^{1,k,l}) \leq \frac{C}{i}.$$

Thus,

$$d_{TV}\left(W_{n,i}^m, NB\left(m, \frac{m}{\mathrm{E}W_{n,i}^m + m}\right)\right) \leq 2e(m+1)\mathrm{P}(X^{*m} \neq W_{n,i}^m)$$
$$\leq 2e(m+1)\left(\mathrm{P}(W_{n,i}^{m,K,L} \neq W_{n,i}) + \mathrm{P}(B_{K,L}^c)\right) \leq \frac{C_m}{i}$$

This proves the statement. ∎

We are now ready to prove three statements that will lead to the proof of our theorem.

Lemma 6.10 *If $W_{n,i}^m$ is the in-degree of vertex i in G_n^m and I is uniform on $\{1, ..., n\}$ independent of $W_{n,i}^m$, then*

1. *Firstly,*
$$d_{TV}\left(W_{n,I}, NB\left(m, \frac{m}{\mathrm{E}W_{n,I}^m + m}\right)\right) \leq C_m \frac{\log n}{n}$$

2. *Secondly,*
$$d_{TV}\left(NB\left(m, \frac{m}{\mathrm{E}W_{n,I}^m + m}\right), NB\left(m, \sqrt{\frac{I}{n}}\right)\right) \leq C_m \frac{\log n}{n}$$

3. *Thirdly,*
$$d_{TV}\left(NB\left(m, \sqrt{\frac{I}{n}}\right), K(m,0)\right) \leq C_m \frac{\log n}{n}$$

Proof.

1. We know from Theorem 6.7 that
$$d_{TV}\left(W_{n,i}^m, NB\left(m, \frac{m}{\mathrm{E}W_{n,i}^m + m}\right)\right) \leq \frac{C_m}{i}.$$

   Thus, applying Lemma 5.6, we find that
$$d_{TV}\left(W_{n,I}, NB\left(m, \frac{m}{\mathrm{E}W_{n,I}^m + m}\right)\right) \leq \mathrm{E}d_{TV}\left(W_{n,I}|I, NB\left(m, \frac{m}{\mathrm{E}W_{n,I}^m + m}\right)|I\right)$$
$$= \sum_{i=1}^n \frac{1}{n} d_{TV}\left(W_{n,i}^m, NB\left(m, \frac{m}{\mathrm{E}W_{n,i}^m + m}\right)\right)$$
$$\leq \sum_{i=1}^n \frac{1}{n} C_m i \leq \frac{C \log n}{n}$$

   This proves the first statement.

2. We first use Lemma 6.6 and Lemma 6.5 to obtain
$$d_{TV}\left(NB\left(m, \frac{m}{\mathrm{E}W_{n,i}^m + m}\right), NB\left(m, \sqrt{\frac{i}{n}}\right)\right) \leq \frac{\frac{m^2}{\mathrm{E}W_{n,i}^m + m} - m\sqrt{\frac{i}{n}}}{\sqrt{\frac{i}{n}}} \leq \frac{C_m}{i}$$

Then, by applying Lemma 5.6, we obtain

$$\begin{aligned}
&d_{TV}\left(NB\left(m,\frac{m}{\mathrm{E}W^m_{n,I}+m}\right),NB\left(m,\sqrt{\frac{I}{n}}\right)\right)\\
&\leq \mathrm{E}d_{TV}\left(NB\left(m,\frac{m}{\mathrm{E}W^m_{n,I}+m}\right)|I,NB\left(m,\sqrt{\frac{I}{n}}\right)|I\right)\\
&=\sum_{i=1}^{n}\frac{1}{n}d_{TV}\left(NB\left(m,\frac{m}{\mathrm{E}W^m_{n,i}+m}\right),NB\left(m,\sqrt{\frac{i}{n}}\right)\right)\\
&\leq \sum_{i=1}^{n}\frac{1}{n}\frac{C_m}{i}\leq\frac{C_m\log n}{n}
\end{aligned}$$

This proves the second statement.

3. Remember that $K(m,0)\stackrel{\mathrm{d}}{=}NB(m,\sqrt{U})$ To prove this item, we couple U to I by writing $U=I/n-V$, with V uniform on $(0,1/n)$ and independent of I. Then,

$$\begin{aligned}
d_{TV}\left(NB\left(m,\sqrt{\frac{I}{n}}\right),K(m,0)\right)&=d_{TV}\left(NB\left(m,\sqrt{\frac{I}{n}}\right),NB(m,\sqrt{I/n-V})\right)\\
&\leq \mathrm{E}d_{TV}\left(\left(NB\left(m,\sqrt{\frac{I}{n}}\right)|I,V\right),\left(NB(m,\sqrt{I/n-V})\right)|I,V\right)\\
&=\sum_{i=1}^{n}\frac{1}{n}\int_0^{1/n}nd_{TV}\left(NB\left(m,\sqrt{\frac{i}{n}}\right),NB(m,\sqrt{i/n-v})\right)dv\\
&\leq\sum_{i=1}^{n}\int_0^{1/n}\frac{m\sqrt{\frac{i}{n}}-m\sqrt{\frac{i}{n}-v}}{\sqrt{\frac{i}{n}}}dv\\
&\leq\sum_{i=1}^{n}\frac{1}{n}\frac{m\sqrt{\frac{i}{n}}-m\sqrt{\frac{i}{n}-\frac{1}{n}}}{\sqrt{\frac{i}{n}}}\\
&\leq\sum_{i=1}^{n}\frac{1}{n}\frac{m\sqrt{\frac{1}{ni}}}{\sqrt{\frac{i}{n}}}=\frac{1}{n}\sum_{i=1}^{n}\frac{m}{i}\leq\frac{C\log n}{n}
\end{aligned}$$

Furthermore, the fifth line follows from the fact that the term within the integral has its maximum at $v=1/n$ and the interval of integration has length $1/n$. We also use the property that $\sqrt{a}-\sqrt{a-1}\leq\sqrt{\frac{1}{a}}$. This shows the statement.

∎

We are now ready to prove the most important result of this section, namely Theorem 6.1. This is an easy proof by the triangle inequality, using the three statements of Lemma 6.10.

Proof.

$$\begin{aligned}
d_{TV}(W_n, K(m,0) \leq & d_{TV}\left(W_{n,I}, NB\left(m, \frac{m}{\mathrm{E}W_{n,I}^m + m}\right)\right) \\
& + d_{TV}\left(NB\left(m, \frac{m}{\mathrm{E}W_{n,I}^m + m}\right), NB\left(m, \sqrt{\frac{I}{n}}\right)\right) \\
& + d_{TV}\left(NB\left(m, \sqrt{\frac{I}{n}}\right), K(m,0)\right) \\
\leq & C_m \frac{\log n}{n}
\end{aligned}$$

This proves the statement. ∎

By proving Theorem 6.1 we have proved the striking power-law behaviour of the random variable describing the in-degree of a randomly selected node from a preferential attachment graph that was explained in Subsection 6.1.

7 Concluding remarks

In this thesis, a very powerful result on a family of preferential attachment graphs is obtained. With Theorem 6.1, we have shown that the in-degree of a randomly selected vertex from a preferential attachment graph approaches the mixed negative binomial distribution defined in Definition 3.4. Furthermore, since the power of Stein's method also lies in quantification of the error, we have shown that the total variation distance between the distribution of the in-degree of a randomly selected vertex from a graph with n vertices and the mixed negative binomial distribution is proportional to $\log n/n$. Also, we have shown that the mixed negative binomial distribution behaves according to a power-law, so these results combined imply that preferential attachment graphs also behave according to a power law. I.e., if picking a random node from a preferential attachment graph, the probability that it has a certain degree decreases exponentially as the degree increases.
This is a striking feature, since this power-law behaviour was not an aim in the construction of the preferential attachment graph, and thus it has great explanatory value. Power-law behaviour in real-world dynamical networks of which the behaviour is similar to preferential attachment models can thus be justified and explained by the properties of preferential attachment graphs.
In [Ross, 2013], a larger family of preferential attachment graphs, with one more parameter, is considered. However, in this thesis this parameter was set 0, to prevent defining the negative binomial distribution with two non-integer parameters. This choice was made to make the calculations and the definition of the negative binomial distribution more intuitive.
To conclude, we have shown how Stein's method can be used to obtain powerful results within probability and in particular to show properties of random graphs. However, the possibilities that Stein's method yield are so much more extensive and this makes it an impressive tool in probability and statistics.

Acknowledgments

I would like to thank my first supervisor Daniel Rodrigues Valesin for the fruitful discussions, enthusiasm and support in writing this thesis. I had no idea that probability could be so much

fun and you were the one to infect me with your passion for the topic. Your commitment was very valuable and I really appreciate that!
Also my second supervisor, Aernout van Enter, was of great help in the process. Your critical eye made me aware of errors that I otherwise would have missed and I would like to thank you for that.

References

[Barbour et al., 1992] Barbour, A. D., Holst, L., and Janson, S. (1992). *Poisson approximation.* Clarendon Press,, Oxford [etc.] :.

[Bollobás et al., 2001] Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290.

[Brown and Phillips, 1999] Brown, T. C. and Phillips, M. J. (1999). Negative binomial approximation with Stein's method. *Methodology And Computing In Applied Probability*, 1(4):407–421.

[Chen et al., 2010] Chen, L. H., Goldstein, L., and Shao, Q.-M. (2010). *Normal approximation by Stein's method.* Springer Science & Business Media.

[Peköz et al., 2013] Peköz, E. A., Röllin, A., and Ross, N. (2013). Total variation error bounds for geometric approximation. *Bernoulli*, 19(2):610–632.

[Pitman and Ross, 2012] Pitman, J. and Ross, N. (2012). Archimedes, Gauss, and Stein. *Notices of the American Mathematical Society*, 59(10):1416–1421.

[Reinert, 2011] Reinert, G. (2011). A short introduction to Stein's method. *Lecture notes, University of Oxford.*

[Ross, 2011] Ross, N. (2011). Fundamentals of Stein's method. *Probability Surveys*, 8(1):210–293.

[Ross, 2013] Ross, N. (2013). Power laws in preferential attachment graphs and Stein's method for the negative binomial distribution. *Adv.in Appl.Probab.*, 45(3):876–893.

[Van Der Hofstad, 2009] Van Der Hofstad, R. (2009). Random graphs and complex networks. pages 159–163. http://www.win.tue.nl/rhofstad/NotesRGCN.pdf.